UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Engenharia Elétrica e de Computação

LEONARDO BRUSCAGINI DE LIMA

DETECÇÃO DE ANOMALIAS EM TEMPO DE RESPOSTA DE SERVIDORES WEB: UMA ABORDAGEM AUTOMATIZADA PARA APRIMORAR A SEGURANÇA E A EFICIÊNCIA.

CAMPINAS

2023

LEONARDO BRUSCAGINI DE LIMA

DETECÇÃO DE ANOMALIAS EM TEMPO DE RESPOSTA DE SERVIDORES WEB: UMA ABORDAGEM AUTOMATIZADA PARA APRIMORAR A SEGURANÇA E A EFICIÊNCIA.

Dissertação apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de mestre em engenharia elétrica, na área de telecomunicações e telemática.

Supervisor/Orientador: Prof. Dr. Yuzo Iano

Este trabalho corresponde à versão final da dissertação defendida pelo aluno Leonardo Bruscagini de Lima, orientada pelo(a) Prof. Dr. Yuzo Iano.

Assinatura do Orientador

CAMPINAS

2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

Lima, Leonardo Bruscagini, 1987-
L628d Detecção de anomalias em tempo de resposta de servidores web : uma abordagem automatizada para aprimorar a segurança e a eficiência / Leonardo Bruscagini de Lima. – Campinas, SP : [s.n.], 2023.

Orientador: Yuzo Iano.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Servidores da Web. 2. Detecção de anomalias. 3. Gerenciamento da informação. 4. Organização da informação. 5. Teoria da informação. I. Iano, Yuzo, 1950-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Informações Complementares

Título em outro idioma: Anomaly detection in web server response time : an automated approuch to enhance security and efficiency
Palavras-chave em inglês:
Web servers
Anomaly detection
Information management
Organization of information
Information theory
Área de concentração: Telecomunicações e Telemática
Titulação: Mestre em Engenharia Elétrica
Banca examinadora:
Yuzo Iano [Orientador]
Gabriel Gomes de Oliveira
Kelem Christine Pereira Jordão
Data de defesa: 10-10-2023
Programa de Pós-Graduação: Engenharia Elétrica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-6427-3698
- Currículo Lattes do autor: http://lattes.cnpq.br/7308909653589666

COMISSÃO JULGADORA – DISSERTAÇÃO DE MESTRADO

Candidato: Leonardo Bruscagini de Lima RA: 152268

Data da Defesa: 10 de outubro de 2023.

Título da Tese: "DETECÇÃO DE ANOMALIAS EM TEMPO DE RESPOSTA DE SERVIDORES WEB: UMA ABORDAGEM AUTOMATIZADA PARA APRIMORAR A SEGURANÇA E A EFICIÊNCIA".

Prof. Dr. Yuzo Iano (Presidente)

Prof. Dr. Gabriel Gomes de Oliveira (Interno–Titular)

Prof. Drª. Kelem Christine Pereira Jordão (Externo–Titular)

A ata de defesa, com as respectivas assinaturas dos membros da Comissão Julgadora, encontra-se no SIGA (Sistema de Fluxo de Dissertação/Tese) e na Secretaria de Pós-Graduação da Faculdade de Engenharia Elétrica e de Computação.

AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos ao Prof. Dr. Yuzo Iano e ao Prof. Dr. Gabriel Gomes de Oliveira pela orientação, suporte e incentivo ao longo desta jornada de preparação em meu programa de pós-graduação na Faculdade de Engenharia Elétrica da UNICAMP.

Suas orientações foram fundamentais para o meu crescimento acadêmico e profissional. Através de suas experiências, conhecimentos e habilidades como pesquisadores e educadores exemplares, fui inspirado a expandir meus horizontes e aprofundar minha compreensão no campo da Engenharia Elétrica e da Computação.

Além disso, gostaria de estender meu agradecimento a toda a comissão de pós-graduação da Faculdade de Engenharia Elétrica da UNICAMP. O apoio e a dedicação de cada membro desta comissão foram essenciais para tornar possível o desenvolvimento desta pesquisa.

Agradeço também aos colegas de laboratório e de curso, cujas discussões e trocas de conhecimento enriqueceram a minha experiência de pós-graduação. As interações com cada um de vocês foram essenciais para a construção de um ambiente acadêmico colaborativo e enriquecedor.

Que todos os meus agradecimentos expressem a imensa gratidão que sinto em meu coração. Esta conquista não teria sido possível sem o apoio e colaboração de cada um de vocês. Espero que nossos caminhos continuem cruzados, e que juntos possamos continuar contribuindo para o avanço da ciência e da engenharia.

“Aprender é a única coisa de que a mente nunca se cansa,

nunca tem medo e

nunca se arrepende.”

(Leonardo da Vinci)

RESUMO

A detecção de anomalias em séries temporais de tempo de resposta de servidores web é um desafio significativo na área de segurança e monitoramento de sistemas. Nesta dissertação, propomos uma metodologia abrangente para a identificação e análise de comportamentos anômalos nos tempos de resposta dos servidores web.

Iniciamos o estudo com a coleta dos dados de tempo de resposta das requisições e a criação de uma base de dados representativa. Utilizamos a biblioteca NumPy para manipulação eficiente dos dados em formato de array, permitindo uma análise mais aprofundada.

Para entender o comportamento médio do servidor, aplicamos a técnica de média móvel, que nos proporcionou uma visão geral dos padrões de resposta ao longo do tempo. Através da variação do tamanho da janela móvel, investigamos a influência dessa técnica na detecção de anomalias e na identificação de tendências.

Além disso, exploramos o uso do boxplot como uma ferramenta gráfica para visualizar a distribuição dos dados e identificar potenciais outliers e comportamentos atípicos. Comparamos os resultados obtidos por meio do boxplot com os das técnicas de média móvel, buscando uma análise mais abrangente das séries temporais.

A partir das abordagens tradicionais de detecção de anomalias, desenvolvemos três métodos automatizados para a identificação de pontos críticos. O primeiro utiliza a média como referência, o segundo incorpora um fator de ajuste, e o terceiro se baseia na variação entre valores consecutivos. Avaliamos a eficácia de cada método em diferentes cenários, considerando a capacidade de detecção precoce e a redução de falsos positivos.

A validação dos métodos propostos foi realizada utilizando um conjunto de dados de testes representativos. Comparando as abordagens com um conjunto de anomalias previamente definidas, avaliamos a capacidade de detecção e a eficiência na identificação dos comportamentos anômalos.

Por fim, apresentamos as conclusões da dissertação, destacando as principais contribuições, limitações e possíveis direções para pesquisas futuras. A metodologia desenvolvida demonstrou sua eficácia na detecção de anomalias em séries temporais de tempo de resposta de servidores web, oferecendo uma abordagem automatizada para melhorar a segurança e o desempenho desses sistemas críticos.

Palavras-chave: Servidores da web; Detecção de anomalias; Gerenciamento da informação; Organização da informação; Teoria da informação.

ABSTRACT

Anomaly detection in time series of web server response time is a significant challenge in the field of security and system monitoring. In this dissertation, we propose a comprehensive methodology for the identification and analysis of anomalous behaviors in web server response times.

We begin the study with the collection of response time data from server requests and the creation of a representative database. We utilize the NumPy library for efficient data manipulation in array format, enabling a more in-depth analysis.

To understand the server's average behavior, we apply the moving average technique, providing us with an overall view of response patterns over time. By varying the size of the moving window, we investigate the influence of this technique on anomaly detection and trend identification.

Moreover, we explore the use of boxplots as a graphical tool to visualize data distribution and identify potential outliers and atypical behaviors. We compare the results obtained from the boxplot with those of moving average techniques, seeking a more comprehensive analysis of time series data.

From traditional anomaly detection approaches, we develop three automated methods for identifying critical points. The first uses the mean as a reference, the second incorporates an adjustment factor, and the third is based on variation between consecutive values. We evaluate the effectiveness of each method in different scenarios, considering early detection capability and reducing false positives.

Validation of the proposed methods is performed using a representative test dataset. By comparing the approaches with a set of pre-defined anomalies, we assess their detection capability and efficiency in identifying anomalous behaviors.

Finally, we present the dissertation's conclusions, highlighting the main contributions, limitations, and possible directions for future research. The developed methodology demonstrated its effectiveness in detecting anomalies in time series of web server response times, offering an automated approach to enhance the security and performance of these critical systems.

Keywords: Web Servers; Anomaly detection; Information management; Organization of information; Information theory.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE GRÁFICOS

LISTA DE ABREVIATURAS, ACRÔNIMOS E SIGLAS.

TIC - Tecnologias da Informação e Comunicação

TI – Tecnologia da Informação

IoT – Internet of Things

GPS - Global Positioning System

CAN bus - Controller Area Network Data Bus

SCADA - Supervisory Control and Data Acquisition

FEBRABAN - Federação Brasileira de Bancos

DoS - Denial of Service

DDoS – Distributed Denial of Service

UNB - University of New Brunswick

IPEA - Instituto de Pesquisa Econômica Aplicada

FBSP - Fórum Brasileiro de Segurança Pública

IJSN - Instituto Jones dos Santos Neves

SINESP - Sistema Nacional de Informações de Segurança Pública

GBPS - Gigabits Por Segundo

IDS - Intrusion Detection System

IPS - Intrusion Prevention System

CSV - Comma-Separated Values

CSE - Communications Security Establishment

CIC - Canadian Institute for Cybersecurity

OWASP - Open Web Application Security Project

HTML - HyperText Markup Language

ARIMA - Autoregressive Integrated Moving Average

RNN - Recurrent Neural Network

CERT - Computer Emergency Response Team

ISO/IEC - International Organization for Standardization/International Electrotechnical Commission

GDPR - General Data Protection Regulation

HIPAA - Health Insurance Portability and Accountability Act

LGPD – Lei Geral de Proteção de Dados

ANSC - Agência Nacional de Segurança Cibernética

ANATEL - Agência Nacional de Telecomunicações

ANS - Agência Nacional de Saúde Suplementar

PCI DSS - Payment Card Industry Data Security Standard

NIST - National Institute of Standards and Technology

SUMÁRIO

1 INTRODUÇÃO

A segurança da informação é um campo essencial em um mundo cada vez mais digital e interconectado. Ela abrange práticas, políticas e tecnologias que visam proteger os dados e sistemas de uma organização contra ameaças, como ataques cibernéticos, roubo de informações confidenciais e interrupções no funcionamento. A garantia da confidencialidade, integridade e disponibilidade dos dados é fundamental para a confiança e a continuidade dos negócios, tornando a segurança da informação um elemento crítico em empresas, governos e na sociedade em geral.

Ela demanda uma abordagem abrangente, envolvendo desde a conscientização dos colaboradores até a implementação de medidas técnicas avançadas para mitigar riscos e garantir a proteção dos ativos de informação.

Os cinco pilares da segurança da informação são princípios fundamentais que ajudam a garantir a proteção e o gerenciamento adequado dos ativos de informação de uma organização.

Figura 1 - Os cinco pilares da segurança da informação

Fonte: Adaptado de ISO 27001. (ABNT NBR, 2013)

Eles são:

- Integridade:
  A integridade da informação refere-se à precisão e à confiabilidade dos dados e informações. Esse pilar assegura que os dados não tenham sido

alterados de forma não autorizada ou acidental. Garantir a integridade significa manter a consistência e a exatidão dos dados ao longo do tempo, prevenindo qualquer forma de corrupção, modificação indevida ou perda de informações.

- Autenticidade:
  A autenticidade está relacionada à verificação da origem da informação. Ela assegura que a fonte da informação seja legítima e que a identidade do remetente ou do criador dos dados possa ser confirmada. Isso impede a falsificação ou a manipulação de informações por partes não autorizadas, garantindo que a informação seja proveniente de fontes confiáveis.
- Confidencialidade:
  A confidencialidade é a garantia de que as informações sensíveis e restritas sejam acessadas somente por pessoas autorizadas. Isso envolve a proteção contra o acesso não autorizado, a divulgação ou o vazamento de informações confidenciais. Medidas como criptografia, controle de acesso e políticas de segurança são usadas para manter a confidencialidade.
- Disponibilidade:
  A disponibilidade diz respeito à garantia de que as informações e os recursos de tecnologia da informação estejam acessíveis e utilizáveis quando necessário. Isso implica a prevenção de interrupções, falhas ou ataques que possam comprometer a acessibilidade dos sistemas e dos dados. Estratégias de redundância, backups e planos de recuperação de desastres são usados para manter a disponibilidade.
- Não Repúdio:
  O não repúdio é a capacidade de provar que uma ação foi realizada por uma determinada parte, de modo que essa parte não possa negar ter executado essa ação. Isso é particularmente importante em transações eletrônicas e comunicações, onde é necessário ter evidências para resolver disputas legais. A assinatura digital e os registros de auditoria são usados para garantir o não repúdio.

Esses cinco pilares são essenciais para a construção de um sistema de segurança da informação eficaz, que visa proteger os ativos de informação de uma organização contra ameaças e garantir a confiabilidade, a integridade e a disponibilidade das informações críticas.

Neste sentido, uma anomalia refere-se a um desvio ou comportamento incomum que pode indicar uma violação ou ameaça aos cinco pilares da segurança da informação. Aqui está como as anomalias podem ser relacionadas a cada um dos pilares:

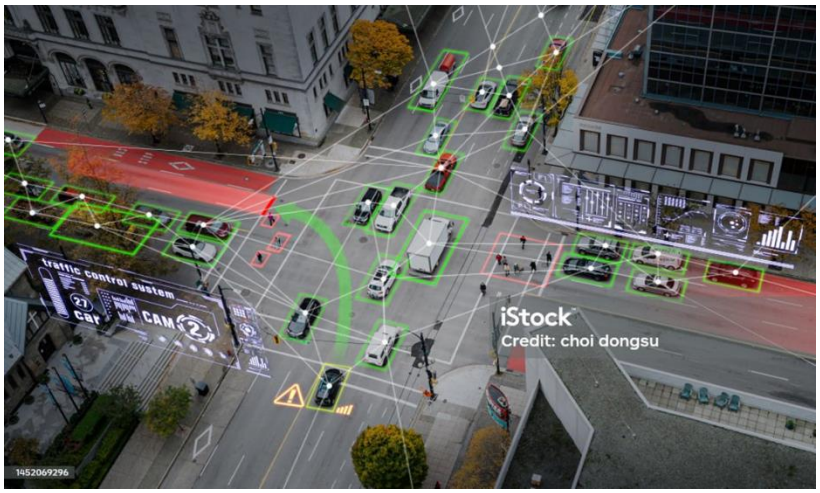
- Integridade: Uma anomalia na integridade dos dados ocorre quando há uma modificação não autorizada ou corrupção dos dados, ameaçando a confiabilidade dos dados.
- Autenticidade: Anomalias na autenticidade podem envolver tentativas de acesso não autorizado a sistemas ou informações por parte de usuários não autenticados, ameaçando a autenticidade dos usuários legítimos.
- Confidencialidade: Anomalias na confidencialidade acontecem quando informações sensíveis estão em risco de exposição a pessoas não autorizadas, colocando em perigo a confidencialidade.
- Disponibilidade: Uma anomalia que afeta a disponibilidade pode incluir ataques de negação de serviço (DoS) que buscam sobrecarregar sistemas e impedir o acesso legítimo, resultando na indisponibilidade de recursos de TI e dados.
- Não Repúdio: Anomalias relacionadas ao não repúdio surgem quando uma parte nega ter realizado uma ação, comprometendo a capacidade de provar a autenticidade de uma transação.

Detectar e responder a anomalias é essencial para a segurança da informação, uma vez que podem indicar atividades maliciosas, erros humanos ou falhas técnicas que afetam os cinco pilares da segurança da informação. Portanto, sistemas de detecção de anomalias desempenham um papel crucial na proteção dos ativos de informação de uma organização.

A detecção e resposta a anomalias no contexto da segurança da informação são áreas em que a transformação digital desempenha um papel crucial. A transformação digital e a conectividade possibilitam a coleta e análise em tempo real

de uma grande quantidade de dados, permitindo uma vigilância mais eficaz em relação aos cinco pilares da segurança da informação.

Figura 2 - Representação de um ecossistema conectado.



Fonte: (PixaBay, 2023)

A transformação digital e a conectividade são conceitos que permeiam diversas áreas da sociedade, desempenhando um papel fundamental na criação de um mundo mais eficiente, inteligente e interconectado. A influência desses conceitos se estende por várias esferas:

Na mobilidade urbana, vemos a integração de tecnologias como aplicativos de transporte e veículos autônomos, bem como sensores em infraestruturas urbanas. Isso se traduz em planejamento de rotas mais eficientes, redução de congestionamentos e promoção de opções de transporte sustentável.

A eficiência e distribuição energética são otimizadas por meio de redes de energia inteligentes, onde medidores e dispositivos estão interligados. Isso possibilita o monitoramento em tempo real, a previsão de demanda e a distribuição de energia de maneira mais eficiente, economizando recursos e reduzindo custos.

Na gestão de resíduos e na coleta de lixo, a transformação digital entra em cena com sensores em contêineres de lixo, permitindo a otimização das rotas de coleta. Isso não apenas reduz custos, mas também contribui para cidades mais limpas e sustentáveis.

Para questões relacionadas ao meio ambiente, a conectividade desempenha um papel crucial na monitorização ambiental. Sensores ajudam a controlar a

qualidade do ar, a poluição da água e a saúde dos ecossistemas, possibilitando uma resposta mais rápida a desastres e uma gestão mais eficaz dos recursos naturais.

A participação das pessoas na democracia participativa é estimulada pela transformação digital, através de plataformas online, aplicativos e mídias sociais que permitem às pessoas se envolverem ativamente na tomada de decisões governamentais, no monitoramento de políticas públicas e na expressão de opiniões.

Em termos de segurança pública, a conectividade é vital, permitindo a interconexão de câmeras de vigilância, sensores de tráfego e sistemas de comunicação. Isso auxilia na prevenção de crimes, resposta mais rápida a incidentes e na gestão de desastres.

Por fim, no âmbito do comércio e da economia, a transformação digital impulsiona o comércio eletrônico e a economia digital. Facilita transações financeiras online, análises de mercado em tempo real, cadeias de suprimentos eficientes e personalização de produtos e serviços, promovendo o crescimento econômico.

Assim, dadas as demandas pelo alto desempenho e aplicação dos conceitos de valor, a cidade incorpora uma nova filosofia: a filosofia do “zero defeito”, tudo deve funcionar perfeitamente bem e sem desperdícios dos recursos de tempo e principalmente financeiros. Para atingir este nível de excelência, dos quais, muito se assemelha com os sistemas de produção das fabricas, faz-se necessário um conhecimento das aplicações e análise dos respectivos históricos de fabricação, bem como das falhas de processos.

Assim como nos processos de fabricação, a escala de progressão da quantidade de dados aconteceu rapidamente e expandiu para as demandas econômicas, sociais até mesmo em termos culturais de entretenimento. As tecnologias da informação e comunicações (TICs) e as técnicas de análise de dados permitem, neste sentido, a possibilidade de cruzamento por meio de diversas técnicas de modo a obter direcionamentos para as tomadas de decisão, permitindo atender à exigência dos consumidores e o aumento da competitividade em todos os mercados.

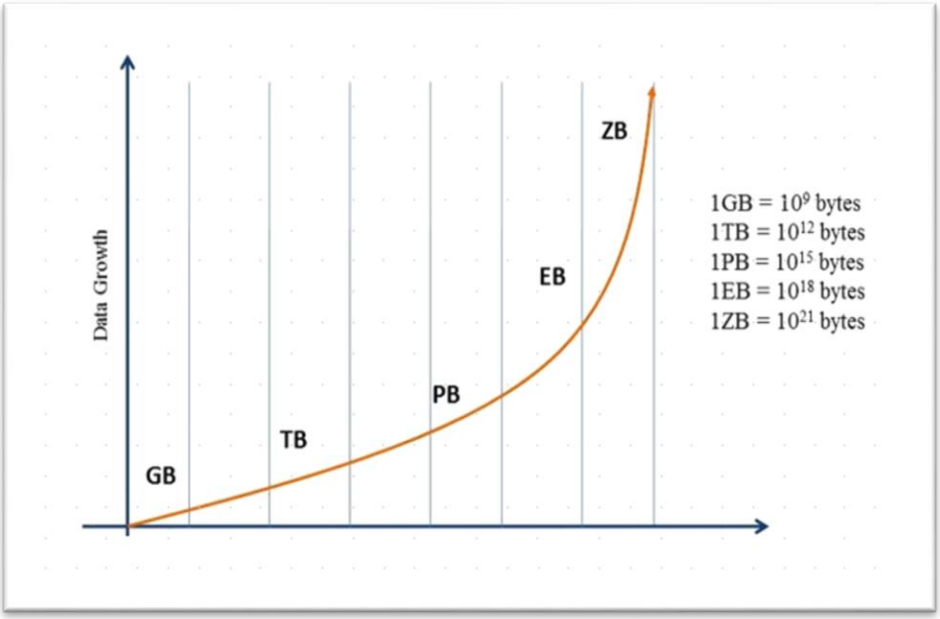
Isso tornou-se possível graças ao armazenamento de grandes volumes de informações, o chamado “Big Data”, e pode ser definido em três pilares (Volume,

Velocidade e Variedade). Trata-se do termo em Tecnologia da Informação (TI) sobre grandes conjuntos de dados que precisam ser processados e armazenados. (RON, 2016)

O gráfico "The Rise of Big Data" é um importante indicador do crescimento da era dos dados. O crescimento do volume de dados está transformando a maneira como as empresas e governos operam.

O gráfico mostra que o volume de dados coletados por empresas e governos está crescendo exponencialmente. Esse crescimento é impulsionado por diversos fatores, incluindo a popularização da tecnologia digital, o aumento da conectividade e o desenvolvimento de novas tecnologias de coleta de dados.

Figura 3: A ascensão dos dados em direção à ‘big data’ ao longo dos anos.



Fonte: (NAGARAJ, SHARVANI, & SRIDHAR, 2018)

O gráfico mostra o crescimento do volume de dados coletados por empresas e governos ao longo de um período de 10 anos, de 2010 a 2020.

A escala de tempo é expressa na parte inferior do gráfico, em um eixo horizontal. O eixo começa em 2010 e termina em 2020. O crescimento do volume de dados é representado por uma linha que se eleva ao longo do tempo.

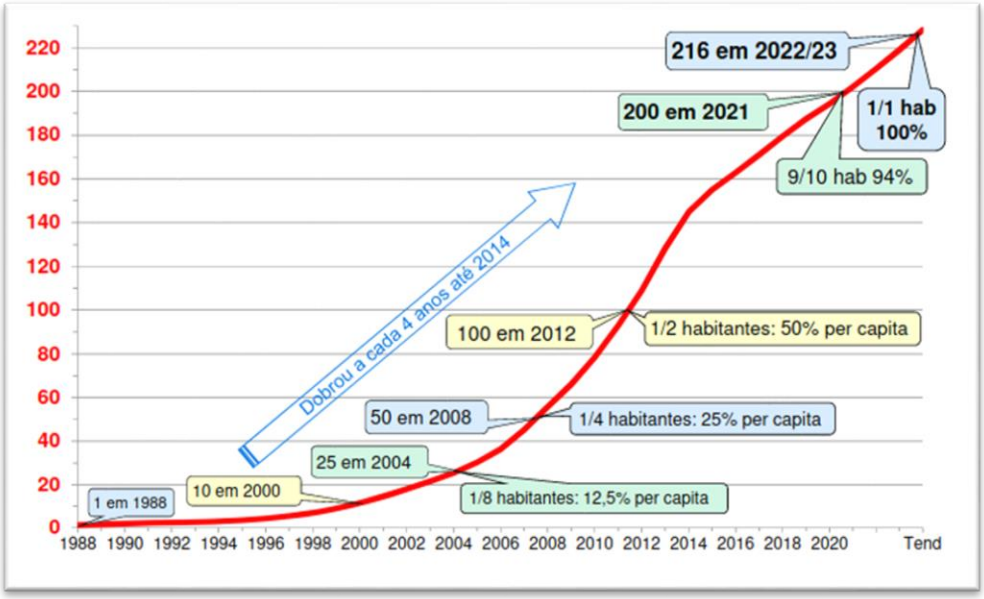
O eixo vertical representa o volume de dados. O eixo começa em 1 zettabyte e termina em 100 zettabytes.

O século XXI, portanto, pode ser considerado como o século da tecnologia da informação e comunicação. Onde os dados trafegam a altas velocidades conectando e processando milhares de grupos e, em consequência, ditando o ritmo da vida moderna. As facilidades e os serviços dos quais permitem a qualidade de vida são tantas que fica difícil imaginar um futuro sem eles. (STEVENSON, 1997)

Afinal de contas, a internet, os dispositivos Internet of Things (IoT), as evoluções dos dispositivos eletrônicos como os processadores e memórias de alto desempenho estão sendo os grandes protagonistas nesta revolução cientifica e tecnológica e por que não, filosófica. Após algumas décadas a humanidade chegou a cerca de 3,4 bilhões de pessoas conectadas à internet em suas casas e um número quatro vezes maior em dispositivos móveis, como os smartphones.

Conforme demonstrou o relatório da 32ª Pesquisa Anual do Uso de TI da FGVcia, mostra que o Brasil possui 216 milhões de computadores (desktop, notebook e tablet) em uso, atingindo um computador por habitante (100% per capita). Esse número representa um aumento de 10% em relação a 2022.

Figura 4: Total de computadores em uso no Brasil (Milhões de unidades – desktops, notebooks e tablets).



Fonte: (MEIRELLES, 2022)

O crescimento do número de computadores em uso no Brasil é impulsionado por diversos fatores, incluindo a popularização da internet, a expansão do ensino a distância e a adoção de tecnologias digitais por empresas e governos.

O relatório também mostra que os smartphones continuam sendo o dispositivo digital mais popular no Brasil, com 242 milhões de unidades em uso. Os tablets, por sua vez, tiveram um crescimento de 20% em 2023, chegando a 42 milhões de unidades.

Em relação às empresas, o relatório mostra que 95% das empresas brasileiras possuem um site. O uso de softwares de gestão empresarial também é crescente, com 65% das empresas brasileiras utilizando esses sistemas.

O relatório da FGVcia é um importante retrato do uso de TI no Brasil. Os dados mostram que o país está se tornando cada vez mais digital, com um crescimento significativo no número de computadores, smartphones e tablets em uso.

O fato de que o Brasil atingiu a marca de um computador por habitante é um marco importante. Isso significa que a população brasileira tem acesso a um dispositivo digital essencial para o trabalho, estudo e entretenimento.

O crescimento do número de computadores em uso é impulsionado pelo aumento da renda e da escolaridade da população brasileira.

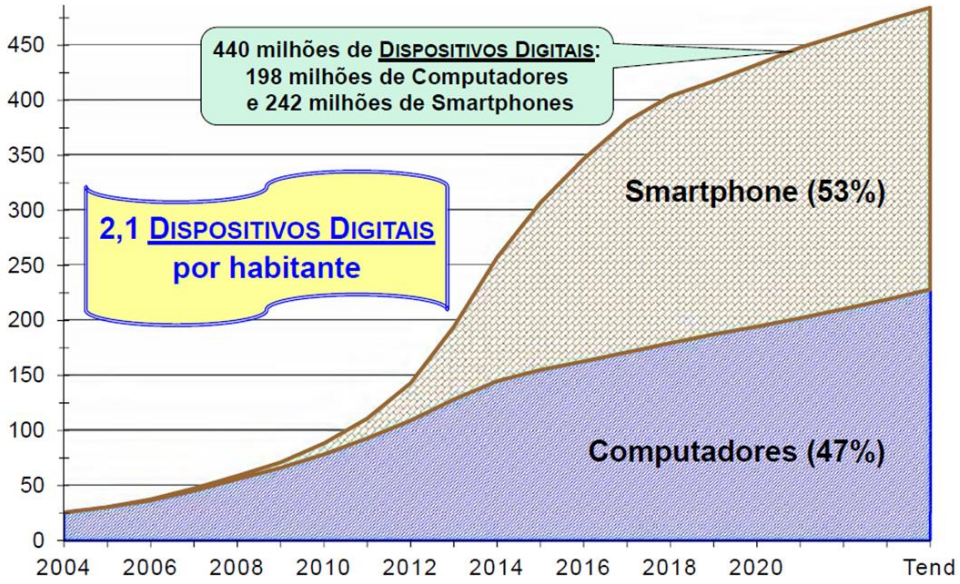
O uso de computadores no Brasil está se tornando mais diversificado. Os tablets, por exemplo, estão se tornando cada vez mais populares, principalmente entre crianças e jovens.

O número de computadores em uso no Brasil cresceu significativamente entre 2010 e 2023. Em 2010, havia 58 milhões de computadores em uso, o que representava 40% da população brasileira. Em 2023, esse número saltou para 215 milhões, o que representa 100% da população brasileira.

O número de smartphones em uso no Brasil também cresceu significativamente entre 2010 e 2023. Em 2010, havia 22 milhões de smartphones em uso, o que representava 15% da população brasileira. Em 2023, esse número saltou para 242 milhões, o que representa 80% da população brasileira.

O número de tablets em uso no Brasil também cresceu significativamente entre 2010 e 2023. Em 2010, havia 1 milhão de tablets em uso, o que representava 1% da população brasileira. Em 2023, esse número saltou para 42 milhões, o que representa 15% da população brasileira.

Figura 5 - Dispositivos digitais (Milhões de unidades)



Fonte: (MEIRELLES, 2022)

No geral, o relatório da FGVcia mostra que o Brasil está se tornando um país cada vez mais digital. O crescimento do número de computadores em uso é um sinal positivo de que a população brasileira está tendo acesso às tecnologias digitais que são essenciais para o desenvolvimento econômico e social.

Estes dispositivos favorecem a coleta massiva de dados, pois compartilham informações com o simples caminhar em uma rua movimentada. Por exemplo, com estas informações é possível medir a pulsação de uma cidade, tornando-as mais agradáveis, inteligentes e abertas. Onde as diferentes tecnologias de comunicação e informação se combinam para criar um ambiente de conectividade. Esta infraestrutura de redes permitiria o monitoramento do trafego de veículos, a qualidade do ar, os indícios de vandalismos chegando as autoridades em tempo real, pois, os dados vão desde a posição do *Global Positioning System* (GPS), até informações do clima, etc.

Este cenário, de uma sociedade inteligente, portanto, poderia elevar a sociedade como uma profunda conhecedora de suas necessidades e tratar problemas crônicos relacionados ao planejamento e ao desperdício. Por exemplo, combater a

desigualdade social, reduzir os índices de criminalidade, elevar taxas de desenvolvimento humano e, caso seja utilizada com responsabilidade, a análise dos dados poderia ser benéfica a todos.

Por outro lado, esta tecnologia possui uma ressalva da mesma magnitude. Os efeitos nocivos de uma inadequada aplicação. As motivações estão associadas ao alcance de vantagem competitiva e obtenção de lucros financeiros. A violação de direitos individuais e da privacidade, controle e monopólio econômico incentivados por pequenos grupos de indústrias de bens de consumos. Regulando, desta maneira, o mercado de ofertas e demandas. Tomando um exemplo, no início dos anos 80, os computadores pessoais ficaram mais acessíveis aos consumidores, e isto levou a um aumento das atividades criminosas. Principalmente no que diz respeito ao auxilio as atividades fraudulentas e acesso indevido de informações. Estas práticas ficaram conhecidas como *cracking*.

Este contexto abre precedente para a exploração de oportunidades que podem ser chamadas vulnerabilidades de sistemas ou expostas em políticas de segurança. A seguir estão alguns dos impactos das vulnerabilidades em redes e sistemas.

Figura 6 - Impacto das vulnerabilidades

Fonte: Adaptado de EC-Council. (EC-Council, 2023)

- Vazamento de informações: Um site ou aplicativo pode expor informações específicas do sistema.

- Negação de serviço: Vulnerabilidades podem impedir que os usuários acessem os serviços de um site ou outros recursos.
- Elevação de privilégio: Atacantes podem obter acesso elevado a um sistema ou recursos protegidos.
- Acesso não autorizado: Atacantes podem obter acesso não autorizado a um sistema, rede, dados ou um aplicativo.
- Roubo de identidade: Atacantes podem ser capazes de roubar informações pessoais ou financeiras dos usuários para cometer fraudes em sua identidade.
- Adulteração de dados: Vulnerabilidades podem levar à recuperação e transmissão não autorizada de dados sensíveis.
- Dano à reputação: Vulnerabilidades podem causar danos à reputação dos produtos e segurança de uma empresa. Danos à reputação têm um impacto direto nos clientes, nas vendas e nos lucros.
- Perda financeira: Danos à reputação podem levar à perda de negócios. Além disso, a exploração de vulnerabilidades pode resultar em despesas para recuperar a infraestrutura de TI danificada.
- Consequências legais: Se os dados pessoais dos clientes forem comprometidos, a organização pode enfrentar consequências legais na forma de multas e sanções regulatórias.

Estes incidentes evidenciaram como os sistemas ligados à Internet são vulneráveis aos ciberataques, pois, a conflitualidade não ocorre com armas físicas, mas via meios eletrônicos e informáticos no ciberespaço.

As invasões são realizadas por organismos, empresas concorrentes ou criminosos de forma a atender interesses específicos. Tendo como principais objetivos a tentativa de expor, alterar, desativar, destruir, roubar ou obter acesso não autorizado ou fazer uso não autorizado de um dispositivo.

Muitos sistemas de governos e de empresas privadas são mantidos na internet e esta é a principal preocupação dos líderes de empresas e de repúblicas, pois os potenciais alvos são as infraestruturas críticas, as redes de energia elétrica, de gás e de água, os serviços de transportes, sistemas supervisórios das plantas industriais, as redes *Controller Area Network* (CAN bus) dos veículos, os módulos IoT, os

sistemas de informação nos mainframes, os serviços de saúde e transações financeiras.

Neste contexto, é importante contar com parceiros no combate as vulnerabilidades. No Brasil, o CERT.br é um importante parceiro para as organizações que desejam melhorar sua segurança contra ataques cibernéticos. Os serviços oferecidos pelo CERT.br ajudam as organizações a identificar, corrigir e mitigar vulnerabilidades, o que reduz o risco de ataques cibernéticos.

As estatísticas do CERT.br são um importante indicador do estado da segurança cibernética no Brasil. Elas mostram que, apesar de uma queda no número de incidentes em 2020, em comparação com anos anteriores, o país ainda é um alvo frequente de ataques cibernéticos.

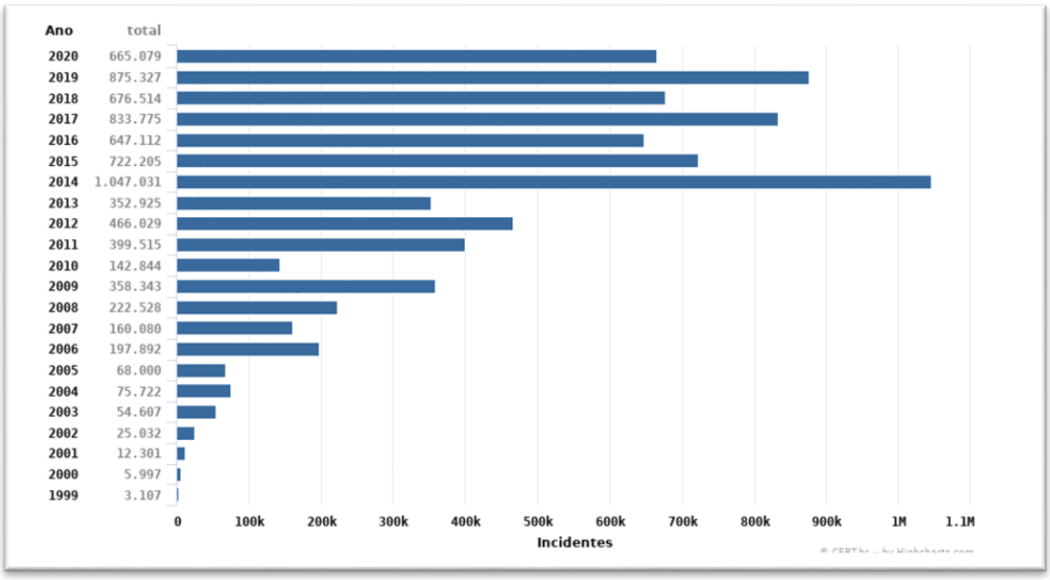
Estas notificações são voluntárias e refletem os incidentes ocorridos em redes que espontaneamente os notificaram ao CERT.br.

Os principais tipos de incidentes reportados ao CERT.br são:

- Ataques de exploração: Esses ataques exploram vulnerabilidades conhecidas em software ou hardware para obter acesso não autorizado a um sistema.
- Ataques de negação de serviço: Esses ataques visam sobrecarregar um sistema com tráfego de rede indesejado, tornando-o indisponível para usuários legítimos.
- Tentativas de fraude: Esses ataques visam enganar as vítimas para que forneçam informações pessoais ou financeiras confidenciais.

E ao analisar as estatísticas do CERT.br, é importante lembrar que elas são apenas uma estimativa do número real de incidentes que ocorrem no Brasil. Muitas empresas e organizações não reportam incidentes ao CERT.br, o que significa que o número real de incidentes pode ser muito maior.

Figura 7: Total de Incidentes reportados ao CERT.br



Fonte: (CERT.br, 2021)

No cenário mundial, as fronteiras físicas se tornam quase irrelevantes, já que as ameaças podem surgir de qualquer lugar do globo. Os ataques virtuais não respeitam limites territoriais e, por isso, exigem uma abordagem abrangente na proteção de sistemas e dados críticos. Todos os países estão suscetíveis a potenciais ataques cibernéticos, independentemente de sua localização geográfica.

A Estônia se tornou um exemplo emblemático da vulnerabilidade aos ataques cibernéticos, quando em 2007 sofreu uma série de ataques que paralisaram empresas e o governo. Embora a Rússia tenha sido acusada, sua origem ainda é desconhecida. A dependência da Internet para serviços no país tornou-o extremamente vulnerável. (BRONZATTI, 2021)

O worm Stuxnet, projetado para atacar o sistema operacional SCADA usados no controle de dispositivos industriais, como centrífugas de enriquecimento de urânio, também se destacou como uma grande ameaça. Isso chamou a atenção para a importância de garantir a segurança dos sistemas eletrônicos e digitais, desde plantas industriais até controladores financeiros.

Os ataques cibernéticos revelaram a vulnerabilidade dos sistemas conectados à Internet e alertaram diversos países sobre a necessidade de proteger seus serviços

virtuais. A ciberguerra emergiu como uma nova forma de conflito global, motivando governos a reforçarem a segurança de seus serviços online. (HOSANG, 2011)

Diante deste cenário desafiador, a cibersegurança emerge como uma prioridade máxima. A colaboração internacional torna-se essencial para compartilhar informações e melhores práticas na proteção contra-ataques cibernéticos. Investimentos em tecnologias avançadas, treinamento e conscientização dos usuários também são fundamentais para fortalecer a defesa contra ameaças virtuais.

Neste sentido, a segurança pública não pode ser tratada apenas como medidas de vigilância e repressão, mas como um sistema integrado e otimizado envolvendo instrumento de prevenção, coação, justiça, defesa dos direitos, saúde e social. O processo de segurança pública se inicia pela prevenção e finda na reparação do dano. (LESSA, 2021)

1.1 Motivação

O cibercrime é uma prática constante no Brasil. As organizações sofrem com estas invasões nos seus dados e informações confidenciais. Os crimes ocasionam não somente danos financeiros, mas também danos empresariais, visto que as organizações têm que fazer novamente a manutenção das máquinas danificadas.

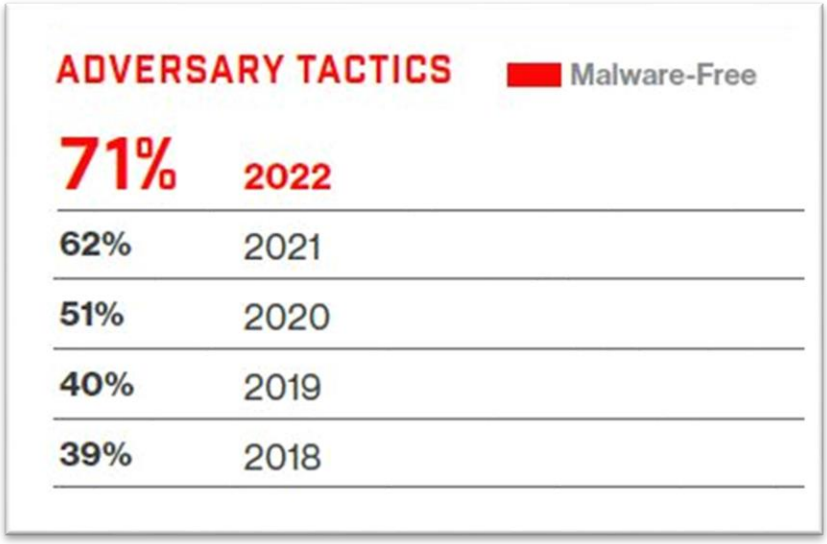
De acordo com relatório da Norton Symantec apontou que 58 milhões de brasileiros foram vítimas de cibercrime em 2021. (Pancini, 2022)

Mas, os criminosos nem sempre utilizam a rede como o objetivo fim. Eles podem, muitas vezes, utilizar a rede como um meio para obter informações das vítimas. Utilizando técnicas de engenharia social. Sim, trata-se de uma técnica empregada por criminosos virtuais para induzir usuários desavisados a enviar dados confidenciais, infectar seus computadores com *malware* ou abrir links para sites infectados. Além disso, os hackers podem tentar explorar a falta de conhecimento do usuário.

Apesar das empresas em geral, governos e sociedade estarem expostos aos riscos da rede, algumas instituições são alvos com maior intensidade. Pois, estão alinhadas aos objetivos dos criminosos. É o caso de bancos e organizações financeiras, por exemplo.

Em 2022, houve uma tendência clara de afastamento do uso de malware, sendo que a atividade sem a presença de malware representou 71% de todas as detecções (um aumento em relação aos 62% de 2021). Isso pode ser parcialmente atribuído à proliferação do abuso de credenciais por parte dos adversários, que buscam facilitar o acesso e a permanência em ambientes das vítimas. Além disso, contribuiu para essa mudança a rapidez com que novas vulnerabilidades foram divulgadas e a agilidade com que os adversários conseguiram operacionalizar *exploits*.

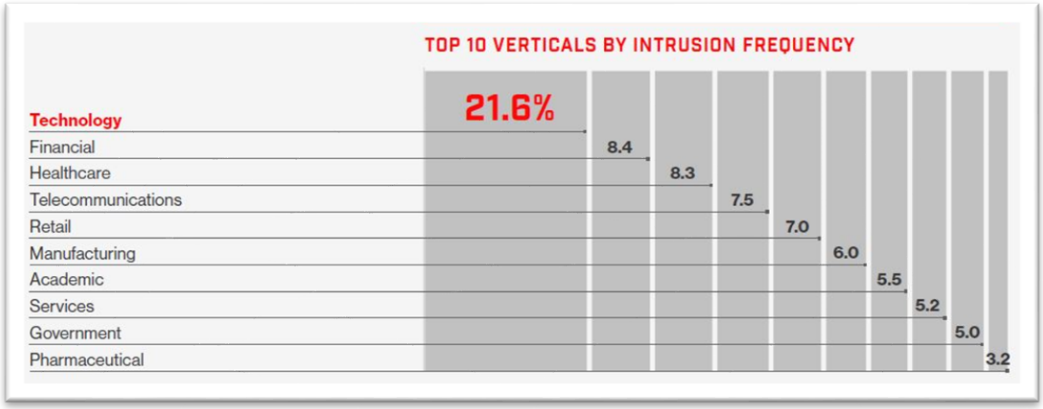
Figura 8: Persistência dos criminosos para obtenção do acesso.



Fonte: (CrowdStrike, 2023)

No quarto trimestre, a CrowdStrike registrou um aumento significativo de 50% no número de campanhas de intrusão interativas com atividade acelerada em comparação com 2021. Além disso, o setor de tecnologia foi o alvo mais frequente dessas invasões interativas em 2022, de acordo com as descobertas do Falcon OverWatch. Essa tendência demonstra um crescimento notável em relação à frequência de invasões nos principais setores da indústria nos 12 meses anteriores.

Figura 9: Comparativo dos setores mais afetados.

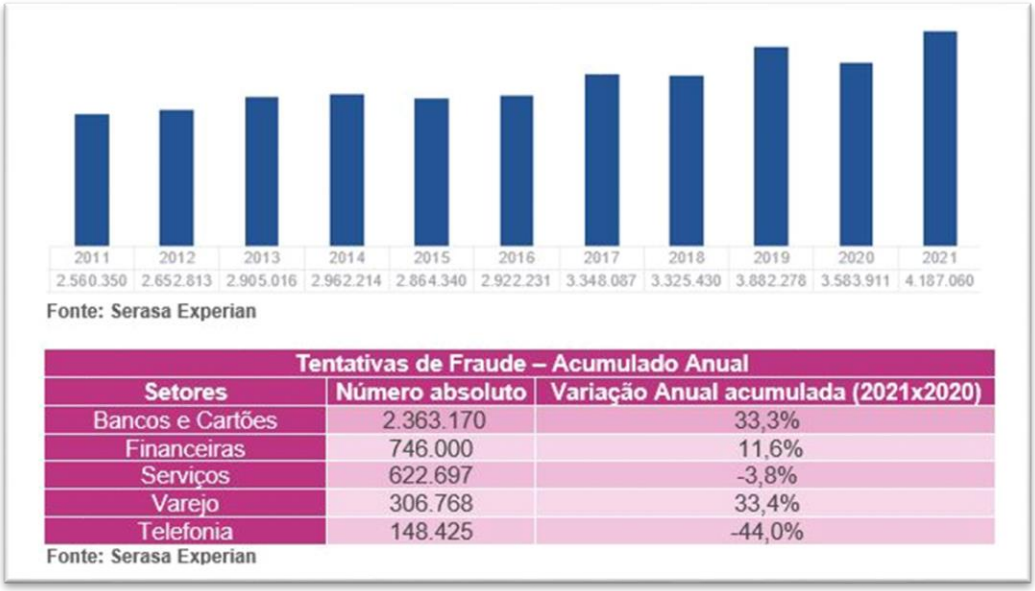


Fonte: (CrowdStrike, 2023)

A Federação Brasileira de Bancos (FEBRABAN) e seus bancos investem constantemente em campanhas e ações de conscientização em seus canais de comunicação com os clientes para orientar a população a se prevenir de fraudes, como uma maneira de atuar no combate e prevenção as fraudes, mas ainda há muito o que se fazer. (CARDOZO, 2019)

O Indicador de Tentativas de Fraude da Serasa Experian registrou no ano de 2021 um total de 4,1 milhões de atividades suspeitas no Brasil. Este valor evidencia um aumento de 16,8% em relação ao acumulado de 2020, quando o índice registrou 3,5 milhões de ataques. É digno de nota que os dados do último ano se destacam por serem os mais substanciais desde o início da série histórica em 2011. (EXPERIAN, 2021)

Figura 10: Tentativas de fraude - acumulado anual



Tentativas de Fraude – Acumulado Anual		
Setores	**Número absoluto**	**Variação Anual acumulada (2021x2020)**
Bancos e Cartões	2.363.170	33,3%
Financeiras	746.000	11,6%
Serviços	622.697	-3,8%
Varejo	306.768	33,4%
Telefonia	148.425	-44,0%

Fonte: Serasa Experian

Fonte: (Experian, 2021)

O enfoque principal dos fraudadores recaiu sobre o segmento de Bancos e Cartões, que assinalou 2,3 milhões de tentativas de fraude, um número também recorde e que representa um crescimento de 33,3% em comparação com o acumulado anual do ano anterior. Em contrapartida, apenas os setores de Serviços e Telefonia experimentaram declínios, com -3,8% e -44,0% respectivamente, nessa comparação.

De acordo com o Diretor de Soluções de Identidade e Prevenção a Fraudes da Serasa Experian, embora os avanços tecnológicos impulsionados pela pandemia sejam, em sua maioria, positivos, também trazem implicações para os consumidores. Ele ressaltou que "o aumento das transações online em 2021 e os diversos novos serviços que passaram a ser oferecidos digitalmente são um terreno fértil para os golpistas. Assim, uma tentativa de fraude ocorre no país a cada 7 segundos. Dessa forma, é essencial que a população redobre a atenção quanto aos dados compartilhados, aos websites acessados e às novas modalidades de golpes que emergem diariamente. A prevenção e a salvaguarda das informações pessoais permanecem como os métodos mais eficazes para evitar fraudes".

Analisando por faixa etária, os consumidores na faixa de 36 a 50 anos foram os mais afetados pelas tentativas de fraude no acumulado anual de 2021, totalizando

1,5 milhão de casos. Aqueles com idades entre 26 e 35 anos foram alvo de 1,1 milhão de ataques. Nas posições subsequentes, estão as pessoas com idades entre 51 e 60 anos (587.791), os mais jovens com até 25 anos (477.512) e, por fim, a parcela da população com mais de 60 anos (461.737).

Ao examinar a distribuição geográfica, o destaque recai sobre a região Sudeste, que registrou 2,1 milhões de atividades suspeitas. A sequência, em ordem decrescente, é composta por Nordeste (726.761), Sul (651.416), Centro-Oeste (375.005) e Norte (250.691). (EXPERIAN, 2021)

No Brasil, a coleta e análise de dados relacionados à segurança pública e à violência desempenham um papel crucial na compreensão e no enfrentamento dos desafios em questões de segurança. Diversas instituições e órgãos governamentais têm trabalhado em conjunto para coletar, processar e apresentar informações relevantes sobre o cenário da segurança no país.

O Ministério da Justiça e Segurança Pública é responsável por coordenar e implementar políticas e ações relacionadas à segurança em âmbito nacional. Para embasar essas ações, o ministério utiliza a plataforma Sinesp (Sistema Nacional de Informações de Segurança Pública), que é uma ferramenta que reúne dados e informações de diferentes órgãos e instituições de segurança, possibilitando a análise de indicadores e a tomada de decisões embasadas em dados concretos.

Além disso, o Instituto de Pesquisa Econômica Aplicada (Ipea) e o Fórum Brasileiro de Segurança Pública (FBSP) têm desempenhado um papel fundamental na pesquisa e análise da violência no Brasil. O Ipea é uma instituição pública que produz estudos e pesquisas voltados para o desenvolvimento econômico e social do país, incluindo análises sobre a segurança pública. O FBSP é uma organização da sociedade civil que busca promover o debate e a reflexão sobre segurança, produzindo relatórios anuais e análises detalhadas sobre a situação da violência no Brasil.

Para enriquecer ainda mais a compreensão do cenário de segurança e violência, o Instituto Jones dos Santos Neves (IJSN) também se une a essa empreitada, colaborando na análise e interpretação dos dados disponíveis. Essa colaboração entre instituições ajuda a construir uma imagem mais completa e precisa dos desafios de segurança que o país enfrenta.

No entanto, é importante mencionar que a coleta e análise de dados de segurança pública podem ser complexas devido a diversos fatores, como subnotificação, falta de padronização nas informações e desafios logísticos. Portanto, essas instituições trabalham constantemente para aprimorar os métodos de coleta e análise, visando a fornecer informações mais precisas e úteis para a formulação de políticas de segurança eficazes. (MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA, 2021)

Neste sentido, a salvaguarda das informações digitais desempenha um papel vital na preservação da segurança nacional e na estabilidade econômica. À medida que a coleta e o compartilhamento de dados digitais se intensificam, a proteção dessas informações torna-se ainda mais crucial tanto para a segurança do país como para sua estabilidade econômica.

A segurança da informação implica em um esforço contínuo para assegurar a integridade de sistemas em rede e todos os dados contra acessos não autorizados ou potencialmente prejudiciais. No âmbito estadual, a segurança nacional e o bem-estar dos cidadãos estão intrinsecamente em jogo. Dessa forma, a consolidação das competências em dados e privacidade, combinando tecnologias disponíveis, pode desempenhar um papel fundamental na aprimorada implementação, operação e manutenção de programas de conformidade e privacidade nas organizações.

A atualização constante dos métodos de gestão de segurança da informação e da infraestrutura revela-se como uma necessidade essencial em uma sociedade livre, onde a garantia dos direitos de privacidade é um pilar indispensável. (BRUSCAGINI, Data Security, Privacy, and Regulatory Issues: A Conceptual Approach to Digital Transformation to Smart Cities. , 2021)

1.2 Objetivo geral

Analisar a natureza e as implicações do cibercrime no contexto global, com foco dos registros de invasões e violações de sistemas, técnicas de invasão a organizações e nos danos financeiros e empresariais resultantes dessas ações no mundo e no Brasil. Tendo em vista, que as corporações possuem negócios globais e interligados, compartilhando tecnologias. Pretende-se compreender as diversas facetas do cibercrime e sua relação com técnicas de engenharia social, bem como investigar os setores mais vulneráveis, como instituições de capital privado e

governos, visando contribuir para a formulação de estratégias de prevenção e combate eficazes.

1.3 Objetivo específico

Analisar o panorama atual do cibercrime no Brasil, incluindo a frequência e os tipos de invasões cibernéticas enfrentadas pelas organizações e seus impactos financeiros e empresariais;

Investigar as técnicas de engenharia social empregadas pelos criminosos virtuais, identificando suas principais estratégias para obter informações confidenciais das vítimas e disseminar malware;

Avaliar a relação entre a falta de conhecimento dos usuários e a exploração dessa vulnerabilidade pelos hackers, destacando os principais pontos de fragilidade na segurança cibernética;

Analisar o papel das instituições financeiras, consultorias privadas, governos, na conscientização e prevenção do cibercrime, examinando as campanhas e ações realizadas para orientar a população e seus impactos na redução das fraudes;

Identificar as principais lacunas e desafios existentes na prevenção e combate ao cibercrime no contexto brasileiro, propondo diretrizes e recomendações para fortalecer as medidas de segurança e minimizar os impactos financeiros e empresariais decorrentes desses ataques.

1.4 Referencial de pesquisa

A estratégia para a condução desta revisão sistemática foi delineada com base nas seguintes considerações:

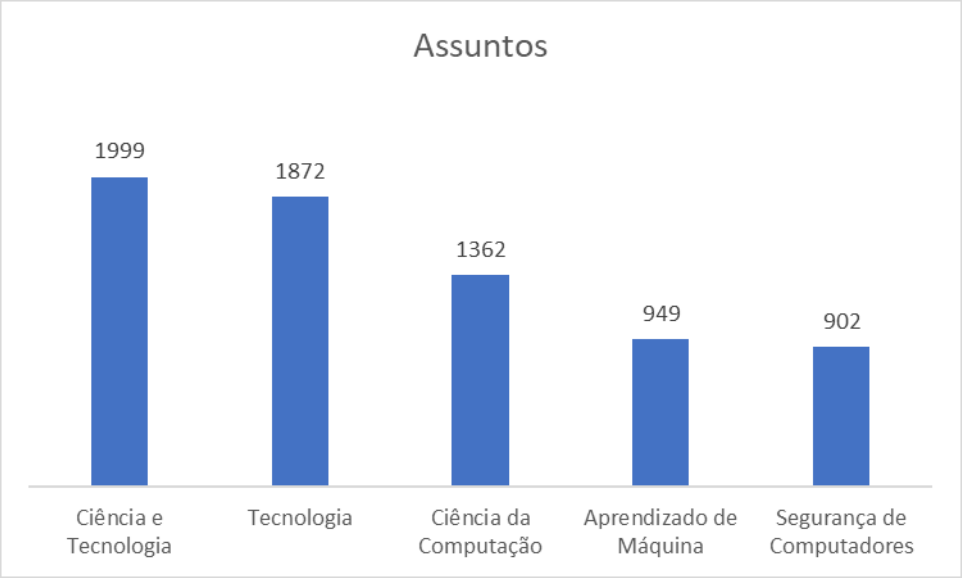
1. Utilização de relatórios provenientes de empresas reconhecidas no campo do monitoramento de fraudes digitais, com o objetivo de coletar dados atuais e confiáveis sobre as tendências e os padrões de atividades fraudulentas no ambiente digital;
2. Incorporação de trabalhos e artigos que fornecem suporte ao contexto social e econômico global e no cenário brasileiro;
3. Busca por relatórios de agências governamentais e empresas do setor privado no Brasil que evidenciem as ameaças em termos quantitativos.

Para ampliar a amplitude e profundidade da pesquisa, adotou-se a utilização do portal de busca integrada Web of Science com as seguintes palavras-chave: "Segurança de servidores web", "Detecção de anomalias em servidores" e "Sistemas de detecção de intrusão (IDS)", todas relacionadas à temática da pesquisa. Isso possibilitou a busca e seleção de artigos científicos, relatórios técnicos e outras fontes pertinentes que contribuirão para uma compreensão mais aprofundada do cenário de fraudes digitais e das estratégias de detecção e prevenção.

O período de análise abrangeu os anos de 2012 a 2022, totalizando 10 anos, resultando em 7084 resultados iniciais.

Os resultados foram os seguintes:

Gráfico 1: Assuntos da pesquisa de referencial teórico



Fonte: (Bruscagini, 2023)

A pesquisa revelou um panorama interessante sobre as preferências e interesses atuais no campo da ciência e tecnologia, com destaque para várias áreas que têm impacto significativo em nossa sociedade em constante evolução.

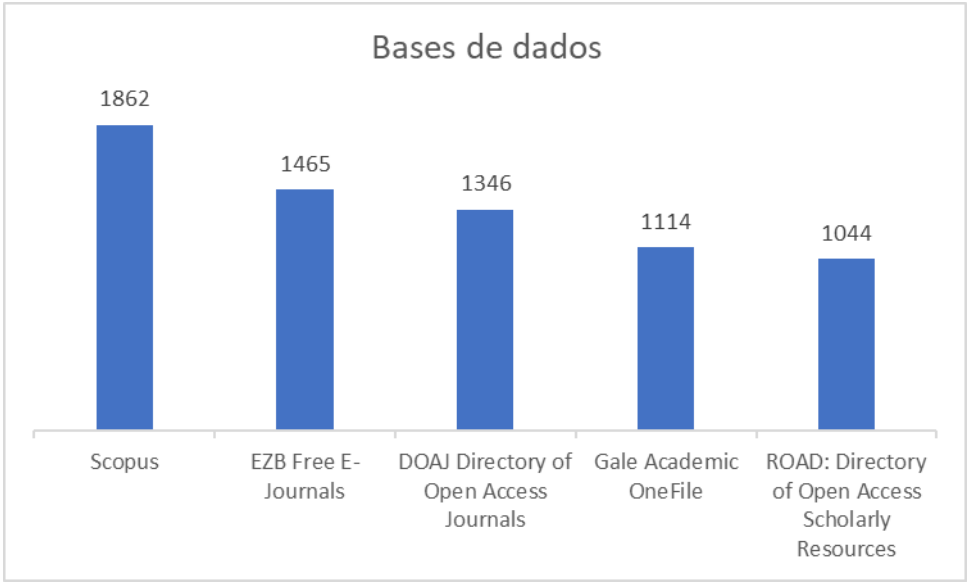
O tópico "Ciência e Tecnologia" aparece como o assunto mais amplo e abrangente, sugerindo um interesse contínuo em entender e explorar os avanços científicos e suas aplicações práticas. Isso pode indicar uma busca por conhecimento interdisciplinar que abarca diversas áreas do saber.

A menção frequente do tópico "Tecnologia" mostra a relevância constante que a inovação tecnológica tem na atualidade. Este é um reflexo claro do papel essencial que a tecnologia desempenha em nossa vida cotidiana, desde a maneira como nos comunicamos até a forma como realizamos transações financeiras e acessamos informações.

O crescente interesse em "Ciência da Computação" e "Aprendizado de Máquina" destaca a importância da computação e da inteligência artificial na nossa era digital. Isso sugere que a exploração das capacidades computacionais e a busca por sistemas autônomos e de aprendizado automatizado estão no centro das atenções.

A inclusão de "Segurança de Computadores" entre os assuntos mais citados é uma indicação clara da crescente preocupação com a proteção dos dados em um mundo cada vez mais digital. Isso reflete os desafios crescentes relacionados à cibersegurança, já que as atividades online tornam-se uma parte integral de nossas vidas.

Gráfico 2: Principais bases de dados e bibliotecas identificados na pesquisa de referencial teórico



Fonte: (Bruscagini, 2023)

Os resultados da pesquisa revelam uma visão esclarecedora sobre as bases de dados e bibliotecas mais utilizadas no contexto da pesquisa acadêmica e científica. Essas plataformas desempenham um papel crucial na disseminação do conhecimento

e no acesso a informações relevantes para a comunidade acadêmica em todo o mundo.

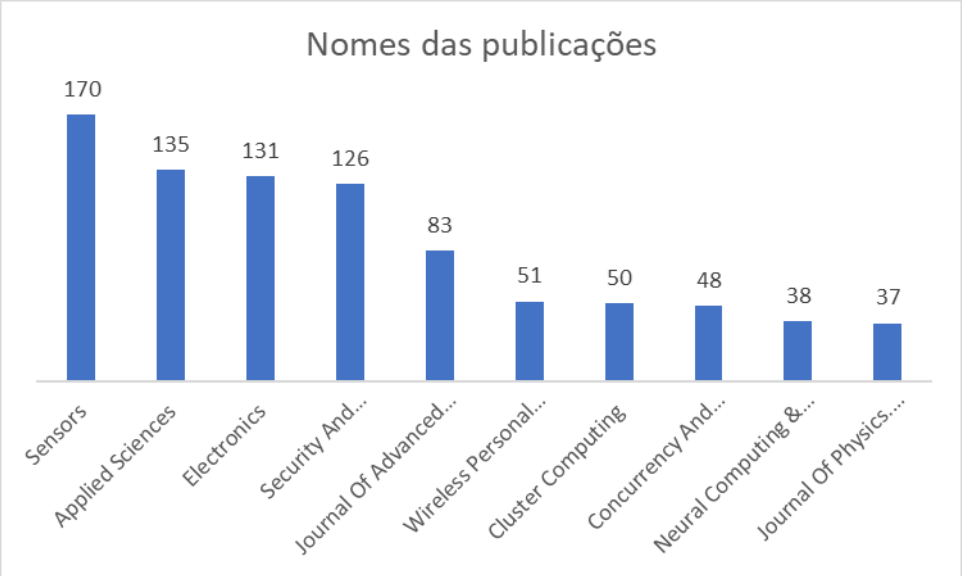
A liderança da plataforma "Scopus" na lista reflete sua posição de destaque como uma das maiores e mais abrangentes bases de dados multidisciplinares. Essa plataforma oferece uma vasta gama de recursos e informações de alta qualidade, permitindo que pesquisadores acessem uma ampla variedade de campos acadêmicos.

A presença de "EZB Free E-Journals" e "DOAJ Directory of Open Access Journals" no topo da lista ressalta a importância crescente do acesso aberto à pesquisa. Essas bases de dados desempenham um papel vital na promoção do acesso livre ao conhecimento, contribuindo para uma maior democratização da informação acadêmica.

"Gale Academic OneFile" também se destaca como uma plataforma amplamente utilizada, oferecendo uma vasta gama de recursos de pesquisa, incluindo artigos de revistas acadêmicas, notícias e outros materiais relevantes. Sua popularidade pode ser atribuída à sua abrangência e facilidade de uso.

A inclusão do "ROAD: Directory of Open Access Scholarly Resources" na lista ressalta ainda mais o foco crescente na pesquisa de acesso aberto. Essa plataforma é uma valiosa fonte para identificar recursos acadêmicos abertos, contribuindo para a colaboração e o compartilhamento de conhecimento entre pesquisadores.

Gráfico 3: Principais publicações identificadas na pesquisa do referencial teórico



Fonte: (Bruscagini, 2023)

A análise dos nomes das publicações revela um panorama interessante das principais revistas científicas e periódicos que são referências em diferentes campos da ciência e tecnologia. Essas publicações desempenham um papel fundamental na disseminação de pesquisas, inovações e descobertas, contribuindo para o avanço do conhecimento em diversas áreas.

"Sensors" se destaca como a publicação líder, com 170 citações. Essa revista desempenha um papel crucial na área de sensores, abordando uma ampla gama de tópicos relacionados a tecnologias de sensoriamento, aplicações e avanços nesse campo em rápida evolução.

"Applied Sciences" e "Electronics" também figuram no topo da lista, com 135 e 131 citações, respectivamente. Ambas as publicações têm foco em diversas disciplinas da ciência aplicada e engenharia eletrônica, fornecendo um espaço para pesquisas inovadoras e práticas.

"Security And Communication Networks" e "Journal Of Advanced Computer Science & Applications" são indicativos da crescente relevância das áreas de segurança cibernética e ciência da computação. Em um mundo cada vez mais digitalizado, a segurança das redes e sistemas de comunicação é de suma importância, e essas publicações refletem a pesquisa ativa nesses campos.

"Wireless Personal Communications", "Cluster Computing" e "Concurrency And Computation" evidenciam a importância das comunicações sem fio, computação em cluster e concorrência de processos na era da tecnologia moderna.

A inclusão de "Neural Computing & Applications" destaca a crescente influência da inteligência artificial e aprendizado de máquina nas pesquisas científicas. A área de redes neurais e suas aplicações diversificadas têm gerado um interesse significativo em diferentes setores.

"Journal Of Physics. Conference Series" fecha a lista, demonstrando a relevância da publicação de trabalhos apresentados em conferências científicas, proporcionando um espaço para compartilhar os resultados de investigações científicas em diversos campos da física.

Os nomes das publicações citadas oferecem um vislumbre das correntes de vanguarda e das evoluções nas distintas esferas da ciência e tecnologia. Não só destacam as áreas onde a pesquisa está dinâmica, mas também ressaltam a crescente interligação entre disciplinas, desempenhando um papel crucial na busca por progresso e saber global.

Para aprimorar ainda mais a precisão dos resultados, uma delimitação temporal foi aplicada, englobando o intervalo de 2017 a 2022, totalizando em 1687 ocorrências. Subsequentemente, critérios adicionais foram incorporados, resultando na exclusão de trabalhos que tratavam de tópicos como tecnologia sem fio, Internet das Coisas, computação em nuvem, tecnologia 5G, sistemas de controle industrial, barramento CAN, blockchain, redes elétricas inteligentes e análise forense digital.

Através destes critérios, 47 trabalhos foram criteriosamente selecionados, sendo dispostos conforme a Tabela 1, de acordo com sua pertinência e relevância. Isso proporciona uma visão mais apurada das áreas de destaque e dos desenvolvimentos de maior importância nos campos de pesquisa considerados.

Tabela 1: Resultado da pesquisa de referencial teórico.

#	Titulo	Autor (es)	Descrição
1	A New Unified Intrusion Anomaly	Kamarudin, Muhammad Hilmi ; Maple, Carsten ; Watson, Tim ;	Esta pesquisa apresenta uma nova abordagem de Sistema de Detecção de Intrusão (IDS) para detectar ataques desconhecidos em servidores web

	Detection in Identifying Unseen Web Attacks	Safa, Nader Sohrabi Del Rey, Ángel Martín ; Ángel Martín Del Rey	usando a abordagem de Detecção de Anomalias de Intrusão Unificada (UIAD).
2	TR-IDS: Anomaly-Based Intrusion Detection through Text-Convolutional Neural Network and Random Forest	Min, Erxue ; Long, Jun ; Liu, Qiang ; Cui, Jianjing ; Chen, Wei Pan, Zhaoqing ; Zhaoqing Pan	Neste artigo, propomos um novo sistema de detecção de intrusão chamado TR-IDS, que aproveita tanto as características estatísticas quanto as características das cargas úteis.
3	A Closer Look at Intrusion Detection System for Web Applications	Agarwal, Nancy ; Hussain, Syed Zeeshan Conti, Vincenzo ; Vincenzo Conti	O objetivo deste artigo de pesquisa é apresentar uma metodologia de design para um IDS eficiente em relação a aplicações web. No artigo, apresentamos vários aspectos específicos que tornam desafiador para um IDS monitorar e detectar ataques web.
4	Analysis of Intrusion Detection Approaches for Network Traffic Anomalies with Comparative Analysis on Botnets	Ahmad, Sultan ; Jha, Sudan ; Alam, Afroj ; Alharbi, Meshal ; Nazeer, Jabeen Azeem, Irshad ; Irshad Azeem	O artigo discute a análise dessas botnets, contramedidas e propõe direções futuras para a pesquisa nessa área. Também explora a análise de energia e futuros estudos sobre a análise de botnets.
5	HADM: detection of HTTP GET flooding attacks by using Analytical hierarchical	Sree, Thankaraja Raja ; Bhanu, Somasundaram Mary Saira	Com o crescimento da Internet, os ataques cibernéticos, incluindo a inundação HTTP GET, se tornaram um problema sério. Detectar esses ataques é difícil, mas o método proposto (HADM) utiliza logs de servidor web, técnicas analíticas e teoria de evidência de Dempster-Shafer para prever

	process and Dempster–Shafer theory with MapReduce		ataques e identificar fontes suspeitas. Os resultados experimentais mostram que o HADM tem alta taxa de detecção, reduz alarmes falsos e é eficiente em termos de tempo de processamento.
6	Cyber Security against Intrusion Detection Using Ensemble-Based Approaches	Alatawi, Mohammed Naif ; Alsubaie, Najah ; Ullah Khan, Habib ; Sadad, Tariq ; Alwageed, Hathal Salamah ; Ali, Shaukat ; Zada, Islam	Este estudo apresenta um método híbrido de seleção de características para melhorar a detecção de intrusões em sistemas IoT. O método alcançou alta precisão em vários tipos de ataques, mostrando sua eficácia na cibersegurança.
7	Towards Effective Detection of Recent DDoS Attacks: A Deep Learning Approach	Ortet Lopes, Ivandro ; Zou, Deqing ; Ruambo, Francis A ; Akbar, Saeed ; Yuan, Bin Li, Wenjuan ; Wenjuan Li	Para abordar os problemas mencionados anteriormente, propomos o CyDDoS, um framework integrado de sistema de detecção de intrusão (IDS), que combina um conjunto de algoritmos de engenharia de recursos com a rede neural profunda. A seleção de recursos em conjunto é baseada em cinco classificadores de aprendizado de máquina usados para identificar e extrair os recursos mais relevantes usados pelo modelo preditivo.
8	A Security Log Analysis Scheme Using Deep Learning Algorithm for IDSs in Social Network	Zhong, Ming ; Zhou, Yajin ; Chen, Gang Peng, Hao ; Hao Peng	Este artigo apresenta um método de análise de registros baseado em aprendizado profundo para um sistema de detecção de intrusões em servidores de redes sociais. Ele descreve os passos para analisar registros, codificar informações de alerta e rastrear a fonte de possíveis intrusões, visando melhorar a segurança do sistema.
9	An Improved Feature Extraction Approach for Web Anomaly Detection Based on	Cheng, Zishuai ; Cui, Baojiang ; Qi, Tao ; Yang, Wenchuan ; Fu, Junsong Liu, Zhe-Li ; Zhe-Li Liu	Este artigo introduz uma abordagem aprimorada de extração de características para firewalls de aplicativos da web baseados em anomalias. A abordagem aproveita a estrutura semântica das URLs para melhorar o desempenho na detecção de ataques web, apresentando uma melhoria média de 5% em métricas de

	Semantic Structure		avaliação em comparação com abordagens convencionais.
10	Design of Intrusion Detection and Prevention in SCADA System for the Detection of Bias Injection Attacks	Benisha, R. B. ; Raja Ratna, S. Gope, Prosanta ; Prosanta Gope	O artigo aborda um sistema de detecção e prevenção de intrusões em sistemas SCADA em tempo real. O sistema utiliza um método de detecção enviesado com criptografia e aprendizado de máquina para detectar e prevenir intrusões, demonstrando maior precisão nos resultados experimentais.
11	A Systematic Review on Hybrid Intrusion Detection System	Maseno, Elijah M. ; Wang, Zenghui ; Xing, Hongyan Maglaras, Leandros ; Leandros Maglaras	Este artigo revisa estudos relacionados a sistemas de detecção de intrusões híbridos entre 2012 e 2022. O artigo destaca a importância e os desafios na criação de sistemas de detecção híbridos eficazes, abordando lacunas existentes e a necessidade de mais pesquisa nessa área.
12	Intrusion detection in mobile ad hoc networks: techniques, systems, and future challenges	Kumar, Sunil ; Dutta, Kamlesh	Este artigo apresenta uma pesquisa estruturada e abrangente das técnicas de detecção de intrusões mais proeminentes para redes móveis ad hoc (MANETs), destacando a importância da detecção em adição às técnicas de prevenção. O artigo classifica as técnicas de detecção em nove categorias e compara suas forças e limitações operacionais, concluindo com sugestões para futuras pesquisas na área de sistemas de detecção de intrusões para MANETs.
13	An Efficient Intrusion Detection Method Based on Federated Transfer Learning and an Extreme Learning	Wang, Kunpeng ; Li, Jingmei ; Wu, Weifei Nazir, Shah ; Shah Nazir	Este estudo propõe o algoritmo para detecção de intrusões, usando aprendizado federado de transferência e máquina de aprendizado extremo. O método proposto alcança melhores resultados de detecção e desempenho robusto, especialmente para amostras pequenas e novas intrusões, protegendo a privacidade dos dados

	Machine with Privacy Preservation		sob o mecanismo de aprendizado federado.
14	Nesting Circles: An Interactive Visualization Paradigm for Network Intrusion Detection System Alerts	Shahryari, Mohammad-Salar ; Mohammad-Khanli, Leyli ; Ramezani, Majid ; Farzinvash, Leili ; Feizi-Derakhshi, Mohammad-Reza Jhaveri, Rutvij ; Rutvij Jhaveri	Este artigo propõe um novo método de visualização de alertas de sistemas de detecção de intrusões chamado "nesting circles", que utiliza círculos para representar informações e fornece uma visualização completa e eficaz de alertas explícitos e implícitos. A eficácia do método é comprovada por meio de um estudo de caso.
15	Preserving Privacy in Multimedia Social Networks Using Machine Learning Anomaly Detection	Aljably, Randa ; Tian, Yuan ; Al-Rodhaan, Mznah Pan, Zhaoqing ; Zhaoqing Pan	Este artigo apresenta um algoritmo de preservação de privacidade que combina técnicas de detecção de anomalias de aprendizado de máquina com modelos de controle de acesso para proteger a privacidade dos usuários em redes sociais multimídia. O método alcançou alta precisão e desempenho superior em comparação com outras técnicas de detecção.
16	An exemplar-based learning approach for detection and classification of malicious network streams in honeynets	Abbasi, Fahim H. ; Harris, Richard ; Marsland, Stephen ; Moretti, Giovanni	Este artigo apresenta um sistema que utiliza aprendizado baseado em exemplos para classificar classes conhecidas de malware e detectar e classificar fluxos maliciosos desconhecidos em classes. Isso é feito criando um modelo usando o menor número possível de exemplos adequados para a classificação de cada classe. Isso permite a detecção e classificação de variantes de malware e novos fluxos maliciosos.
17	Improving the Accuracy of Network Intrusion Detection with Causal Machine Learning	Zeng, Zengri ; Peng, Wei ; Zhao, Baokang Tan, Zhiyuan ; Zhiyuan Tan	Esta pesquisa visa estabelecer um novo sistema de detecção de intrusão em rede (NIDS) baseado em ML causal. O sistema proposto começa com a identificação de características ruidosas por meio de intervenção causal, mantendo apenas as características que têm causalidade com ciberataques. Em seguida, o algoritmo de ML é usado

			para fazer uma classificação preliminar e selecionar os tipos mais relevantes de ciberataques.
18	Empirical Evaluation of Noise Influence on Supervised Machine Learning Algorithms Using Intrusion Detection Datasets	Al-Gethami, Khalid M. ; Al-Akhras, Mousa T. ; Alawairdhi, Mohammed Chen, Tom ; Tom Chen	Este artigo examina empiricamente o efeito do ruído na precisão dos IDSs baseados em ML por meio da realização de um amplo conjunto de experimentos diferentes. Os algoritmos de ML utilizados são árvore de decisão (DT), random forest (RF), máquina de vetor de suporte (SVM), redes neurais artificiais (ANNs) e Naïve Bayes (NB).
19	An Efficient Communication Intrusion Detection Scheme in AMI Combining Feature Dimensionality Reduction and Improved LSTM	Lu, Guanyu ; Tian, Xiuxia Han, Jinguang ; Jinguang Han	Métodos baseados em LSTM têm uma habilidade superior para detectar tráfego anormal, mas não conseguem extrair características estruturais bidirecionais. Projetamos um modelo Bi-directional Long Short-Term Memory (BiLSTM) que adicionou um Mecanismo de Atenção. Ele pode determinar a criticidade da dimensionalidade e melhorar a precisão do modelo de classificação.
20	Recurrent Neural Network Model Based on a New Regularization Technique for Real-Time Intrusion Detection in SDN Environments	Albahar, Marwan Ali Li, Huaizhi ; Huaizhi Li	Neste artigo, propomos um modelo de rede neural recorrente (RNN) baseado em uma nova técnica de regularização (RNN-SDR). Essa técnica suporta a detecção de intrusões em SDNs. O objetivo da regularização é generalizar o modelo de aprendizado de máquina o suficiente para que ele seja executado de forma otimizada.
21	Use of Security	Ávila, Ricardo ; Khoury, Raphaël	Este artigo apresenta uma revisão sistemática da literatura sobre o uso de

	Logs for Data Leak Detection: A Systematic Literature Review	; Khoury, Richard ; Petrillo, Fábio Lombardi, Flavio ; Flavio Lombardi	registros de segurança para detecção de vazamento de dados. Nossas descobertas são quatro: (i) propomos uma nova classificação de vazamentos de informações, que utiliza os princípios do GDPR; (ii) identificamos os vinte conjuntos de dados de ameaças mais amplamente utilizados e publicamente disponíveis; (iii) descrevemos vinte tipos de ataques presentes em conjuntos de dados públicos; e (iv) descrevemos trinta algoritmos utilizados para detecção de vazamento de dados.
22	Network Intrusion Detection with Threat Agent Profiling	Bajtoš, Tomáš ; Gajdoš, Andrej ; Kleinová, Lenka ; Lučivjanská, Katarína ; Sokol, Pavol Díaz-Verdejo, Jesús ; Jesús Díaz-Verdejo	Neste artigo, discutimos abordagens que simplificam o trabalho dos administradores de rede. Aplicamos métodos de agrupamento para a criação de perfis de incidentes de segurança. Consideramos os algoritmos de agrupamento K-means, PAM e CLARA.
23	Black Hole Attack Detection Using K-Nearest Neighbor Algorithm and Reputation Calculation in Mobile Ad Hoc Networks	Farahani, Gholamreza Chaudhry, Shehzad Ashraf ; Shehzad Ashraf Chaudhry	Este artigo propõe um novo algoritmo em MANETs para detectar ataques de buraco negro usando o algoritmo K-vizinhos mais próximos (KNN) para o agrupamento e inferência fuzzy para selecionar o líder do cluster. Com o uso da distribuição beta e lógica mental de Josang, a confiança de cada nó será calculada.
24	Detection and classification of anomaly intrusion using hierarchy clustering and SVM	Tang, Chenghua ; Xiang, Yang ; Wang, Yu ; Qian, Junyan ; Qiang, Baohua	Este artigo apresenta um novo modelo hierárquico de detecção de intrusão por anomalia que combina o Fuzzy c-means com algoritmo genético e o SVM.

25	Network Attacks Detection Methods Based on Deep Learning Techniques: A Survey	Wu, Yirui ; Wei, Dabao ; Feng, Jun Xu, Xiaolong ; Xiaolong Xu	Neste artigo, apresentamos uma revisão sobre métodos de detecção de ataques que envolvem a força de técnicas de deep learning.
26	A Hidden Attack Sequences Detection Method Based on Dynamic Reward Deep Deterministic Policy Gradient	Zhang, Lei ; Pan, Zhisong ; Pan, Yu ; Guo, Shize ; Liu, Yi ; Xia, Shiming ; Zheng, Qibin ; Li, Hongmei ; Bai, Wei Karuppiah, Marimuthu ; Marimuthu Karuppiah	Propusemos um método de detecção de sequências de ataque ocultas baseado em aprendizado por reforço para lidar com esse desafio, modelando os administradores de rede como um agente inteligente que aprende sua política de ação a partir da interação com o ambiente do ciberespaço
27	BLATTA: Early Exploit Detection on Network Traffic with Recurrent Neural Networks	Pratomo, Baskoro A. ; Burnap, Pete ; Theodorakopoulos, George Amjad, Muhammad Faisal ; Muhammad Faisal Amjad	Propomos um novo mecanismo de detecção precoce de exploits que examina o tráfego de rede, lendo apenas 35,21% das mensagens da camada de aplicação para prever o tráfego malicioso, mantendo uma taxa de detecção de 97,57% e uma taxa de falsos positivos de 1,93%.
28	Comparing and Analyzing Applications of Intelligent Techniques in Cyberattack Detection	Dixit, Priyanka ; Kohli, Rashi ; Acevedo-Duque, Angel ; Gonzalez-Diaz, Romel Ramon ; Jhaveri, Rutvij H. Shafiq, Muhammad ; Muhammad Shafiq	Este artigo discute as melhorias e o aprimoramento de modelos de segurança, estruturas para a detecção de ciberataques e prevenção por meio de diferentes técnicas de aprendizado de máquina e otimização no domínio da cibersegurança. O foco está na literatura de diferentes algoritmos metaheurísticos para seleção de recursos ótimos e técnicas de aprendizado de máquina para a classificação de ataques.

29	Traffic classification for managing Applications' networking profiles	Alizadeh, Hassan ; Zúquete, André	Este artigo analisa metodologias de classificação de tráfego, dentro de um framework de taxonomia, para encontrar as melhores metodologias de classificação de tráfego que poderiam nos ajudar a responder à seguinte pergunta: dado uma amostra de tráfego, gerada por uma aplicação específica, ela está em conformidade com o tráfego esperado da aplicação?
30	Cooperative security management for broadband network environments	Cruz, Tiago ; Simões, Paulo ; Monteiro, Edmundo ; Bastos, Fernando ; Laranjeira, Alexandre	Este artigo descreve um sistema de detecção de intrusões distribuído (DIDS) projetado para utilizar gateways residenciais (RGWs) como agentes de segurança ativos para o ambiente doméstico.
31	Detection System of HTTP DDoS Attacks in a Cloud Environment Based on Information Theoretic Entropy and Random Forest	Idhammad, Mohamed ; Afdel, Karim ; Belouch, Mustapha Li, Huaizhi ; Huaizhi Li	Neste artigo, apresentamos um sistema de detecção de ataques HTTP DDoS em um ambiente em nuvem com base na Entropia Teórica da Informação e no algoritmo de aprendizado em conjunto Random Forest. Um algoritmo de janela deslizante baseado no tempo é usado para estimar a entropia das características do cabeçalho da rede do tráfego de entrada da rede.
32	Evaluation of entropy-based detection of outbound denial-of-service attacks in edge networks	Basicevic, Ilija ; Ocovaj, Stanislav ; Popovic, Miroslav	Este artigo apresenta uma avaliação da detecção de intrusões em redes baseada em entropia no caso de ataques de negação de serviço (DoS) de saída em redes de borda. O detector monitora a entropia de várias distribuições simples de pacotes: portas de origem e destino, e número de pacotes e bytes transferidos.
33	Research on DoS Traffic Detection Model Based	He, Hongyan ; Huang, Guoyan ; Zhang, Bing ; Zheng, Zhangqi	O artigo propõe um método eficiente de detecção de tráfego de ataque DoS, o algoritmo híbrido de detecção de ataque

	on Random Forest and Multilayer Perceptron	Arif, Muhammad ; Muhammad Arif	de rede Random Forest e Multilayer Perceptron (RF-MLP).
34	Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning	Lima Filho, Francisco Sales de ; Silveira, Frederico A. F. ; de Medeiros Brito Junior, Agostinho ; Vargas-Solar, Genoveva ; Silveira, Luiz F. Maglaras, Leandros ; Leandros Maglaras	Este artigo apresenta um sistema de detecção de DoS baseado em aprendizado de máquina (ML). A abordagem proposta faz inferências com base em assinaturas previamente extraídas de amostras de tráfego de rede. Os experimentos foram realizados usando quatro conjuntos de dados de referência modernos.
35	Towards reducing false alarms in network intrusion detection systems with data summarizati on technique	Hubballi, Neminath ; Biswas, Santosh ; Nandi, Sukumar	Neste artigo, é proposto um IDS de anomalia capaz de lidar com conjuntos de dados grandes, mas minimizando alarmes falsos. Os resultados experimentais em três conjuntos de dados diferentes ilustram a eficácia desse método para lidar com grandes volumes de dados e aumentar a precisão.
36	A Framework for Real-Time Intrusion Response in Software Defined Networking Using Precompute d Graphical Security Models	Eom, Taehoon ; Hong, Jin B. ; An, SeongMo ; Park, Jong Sou ; Kim, Dong Seong Cimato, Stelvio ; Stelvio Cimato	Neste artigo, propomos uma resposta a intrusões em tempo real em SDN usando pré-computação para estimar a probabilidade de futuros caminhos de ataque a partir de um ataque em curso.
37	A framework for intrusion detection	Beigi Mohammadi, Nasim ; Mišić,	Neste artigo, discutimos os requisitos de segurança e as vulnerabilidades da AMI e revisamos as soluções existentes de

	system in advanced metering infrastructure	Jelena ; Mišić, Vojislav B. ; Khazaei, Hamzeh	prevenção e detecção de ameaças. Propomos um IDS para a rede de área de vizinhança (NAN) na AMI, levando em consideração os requisitos específicos da NAN.
38	HeteMSD: A Big Data Analytics Framework for Targeted Cyber-Attacks Detection Using Heterogeneous Multisource Data	Ju, Ankang ; Guo, Yuanbo ; Ye, Ziwei ; Li, Tao ; Ma, Jing Angin, Pelin ; Pelin Angin	Neste artigo, primeiro revisamos diferentes mecanismos de fusão de dados que correlacionam dados heterogêneos de várias fontes.
39	A test of intrusion alert filtering based on network information	Sommestad, Teodor ; Franke, Ulrik	Este artigo relata um teste realizado para avaliar várias alternativas de filtragem que aproveitam informações sobre as propriedades estáticas da rede de computadores monitorada, como vulnerabilidades e exposição de portas e hosts.
40	Intraclass and interclass correlation coefficient-based feature selection in NIDS dataset	Vasudevan, Alampallam Ramaswamy ; Selvakumar, Subramanian	Neste artigo, é proposta a aplicação do coeficiente de correlação intraclasse e do coeficiente de correlação interclasse para alcançar um subconjunto eficiente de características específicas de classe alvo
41	Improving intrusion detection for imbalanced network traffic	Thomas, Ciza	O objetivo deste trabalho é fornecer uma arquitetura que permita que sistemas de detecção de intrusões disponíveis trabalhem juntos para criar um modelo mais realista do estado de uma rede.
42	A model of analyzing cyber threats	Choi, Sang-soo ; Song, Jungsuk ;	Este artigo apresenta um modelo de monitoramento e resposta de segurança baseado no tráfego da darknet,

	trend and tracing potential attackers based on darknet traffic	Kim, Seokhun ; Kim, Sookyun	composto por seis componentes principais: endereços IP da darknet, sistema de detecção de intrusões, servidor de coleta, sistema de gerenciamento, sistema de análise e sistema de rastreamento. Isso nos permite obter tráfego real que pode conter códigos de ataque e padrões de ataque maliciosos.
43	Network specific vulnerability based alert reduction approach	Njogu, Humphrey Waita ; Jiawei, Luo ; Kiere, Jane Nduta	Este artigo busca abordar o problema mencionado acima para fortalecer as abordagens de gerenciamento de alertas com base em vulnerabilidades. Nossa abordagem verifica os alertas antes de mesclá-los. Central para essa abordagem é o uso de dois componentes: verificador e mesclador de alertas.
44	Feature engineering for detection of Denial of Service attacks in session initiation protocol	Asgharian, Hassan ; Akbari, Ahmad ; Raahemi, Bijan	Neste artigo, estudamos as preocupações de segurança dos sistemas baseados em SIP e propomos um conjunto de recursos para eles.
45	A Smart Agent Design for Cyber Security Based on Honeypot and Machine Learning	El Kamel, Nadiya ; Eddabbah, Mohamed ; Lmoumen, Youssef ; Touahni, Raja Shaukat, Sajjad ; Sajjad Shaukat	Neste artigo, apresentamos uma introdução sobre aprendizado de máquina e sistemas de honeypot, e com base nessas tecnologias, projetamos um agente inteligente para prevenção e previsão de ciberataques.
46	Defeat scanning worms in cyber warfare	Hsu, Fu-Hau ; Chen, Li-Han ; Lin, Chia-Jun	Neste artigo, propomos um sistema de defesa automático chamado Sistema Serum, contra worms de varredura.

47	On the IEEE 802.11i security: a denial-of-service perspective	Singh, Rajeev ; Sharma, Teek Parval	O artigo apresenta uma revisão dos ataques de DoS e das soluções existentes relacionadas ao padrão de segurança IEEE 802.11i.

1.5 Organização

No capítulo de introdução, a abordagem inicial explora a filosofia do "zero defeito" presente nas cidades modernas, enfocando o alto desempenho e a aplicação de valores como elementos centrais. Essa análise visa compreender as motivações de diversos setores da sociedade, visando a otimização de recursos e a maximização da eficiência. Além disso, destaca-se a convergência em curso entre sistemas de produção industrial e demandas urbanas, um processo impulsionado pelas tecnologias de informação e comunicação (TICs).

Nesse contexto de transição para a era da tecnologia da informação e comunicação, é ressaltada a importância das redes de comunicação, dispositivos IoT e tecnologias avançadas. No entanto, essa transformação também traz consigo desafios significativos no que tange à segurança cibernética. O capítulo se aprofunda em uma análise abrangente sobre a natureza e as implicações do cibercrime global, focalizando invasões de sistemas, técnicas empregadas nesses ataques e os impactos financeiros decorrentes. Adicionalmente, explora-se a exploração da lacuna de conhecimento por parte dos usuários e a proposta de diretrizes que visam mitigar e combater o cibercrime, particularmente no contexto brasileiro.

No segundo capítulo, uma revisão bibliográfica é realizada para estabelecer o arcabouço conceitual e metodológico da pesquisa. São discutidos temas como segurança de servidores web, análise de séries temporais para detecção de anomalias, estatísticas de violações de segurança, bem como métodos de prevenção. Normas e regulamentações nacionais e internacionais pertinentes também são abordadas.

O capítulo três se dedica a detalhar a metodologia adotada, explorando abordagens estatísticas e computacionais para detectar anomalias em séries temporais de tempo de resposta de servidores web. Salienta-se a importância da

automatização desse processo para apoiar decisões rápidas e eficientes por parte dos administradores.

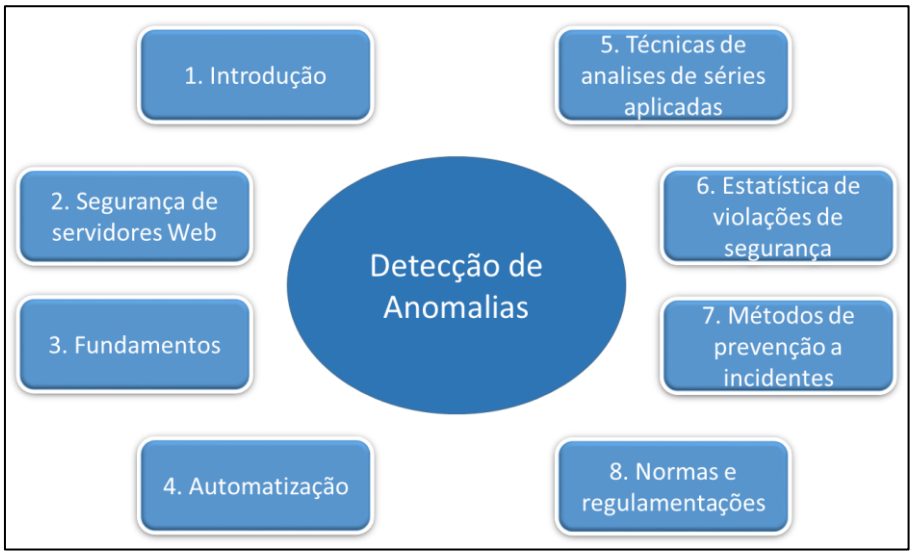
No capítulo seguinte, os resultados da análise são apresentados. Através da aplicação de diferentes métodos de detecção de anomalias em dados de tempo de resposta, são oferecidas perspectivas únicas para identificar comportamentos atípicos, enriquecendo a compreensão das variações e padrões presentes.

O capítulo final encerra o trabalho reunindo os conceitos abordados, realçando como a combinação dos métodos amplia a compreensão dos dados. Há uma ênfase no equilíbrio entre sensibilidade na detecção de anomalias e a minimização de falsos positivos para garantir resultados confiáveis e acionáveis. Por fim, o capítulo seis aponta para futuros direcionamentos da pesquisa, aplicações potenciais e áreas de desenvolvimento subsequente.

2 REVISÃO BIBLIOGRAFICA

Neste capítulo, serão apresentados os principais conceitos e fundamentos que constituem a base teórica da pesquisa sobre detecção de anomalias em tempo de resposta de servidores web. Os tópicos abordados incluirão os conceitos fundamentais de segurança de servidores web, as técnicas de análise de séries temporais aplicadas à detecção de anomalias, estatísticas de violações de segurança da informação e privacidade, métodos e técnicas de prevenção de incidentes de segurança em servidores web, bem como as normas e regulamentações nacionais e internacionais relevantes para a segurança de servidores web e detecção de anomalias.

Figura 11 - Representação dos tópicos dos capítulos da revisão bibliográfica



Fonte: (Bruscagini, 2023)

2.1 Introdução as Detecção de Anomalias

A detecção de anomalias desempenha um papel vital na segurança de servidores web, especialmente em um cenário onde as ameaças cibernéticas estão em constante evolução e tornam-se cada vez mais sofisticadas. Enquanto os sistemas de segurança tradicionais tendem a se concentrar na defesa contra ameaças conhecidas, a detecção de anomalias visa identificar comportamentos e atividades que fogem do padrão estabelecido. Isso é particularmente importante porque muitas ameaças modernas são projetadas para evitar serem detectadas pelos métodos convencionais de segurança. (Eskin, 2002)

A importância da detecção de anomalias é acentuada pelo fato de que as ameaças cibernéticas raramente seguem um único padrão. Os invasores empregam táticas diversificadas, explorando vulnerabilidades em camadas distintas de sistemas e redes. Isso torna difícil confiar somente em assinaturas de ameaças conhecidas ou em soluções de segurança baseadas em regras estáticas. A detecção de anomalias aborda essa lacuna, sendo capaz de identificar atividades suspeitas, mesmo quando os vetores de ataque são novos ou desconhecidos.

2.1.1 Objetivos da detecção de anomalias automatizada, aprendizado e apoio à tomada de decisões

A detecção de anomalias tem vários objetivos cruciais. Primeiramente, visa identificar padrões de comportamento anormal que podem indicar atividades maliciosas ou não autorizadas. Isso inclui não apenas ataques diretos, mas também atividades suspeitas que podem indicar preparativos para futuros ataques.

Além disso, a detecção de anomalias automatizada é uma ferramenta poderosa para aprendizado contínuo. À medida que novos dados são coletados e analisados, os algoritmos de detecção podem ajustar seus modelos para incorporar comportamentos emergentes. Isso é particularmente útil em ambientes onde as ameaças estão sempre mudando, permitindo que os sistemas de segurança acompanhem as últimas tendências. (Antonakakis, 2010)

Um aspecto frequentemente subestimado da detecção de anomalias é seu papel no apoio à tomada de decisões. Ao alertar os administradores de sistemas e equipes de segurança sobre atividades suspeitas, a detecção de anomalias fornece informações valiosas que podem ser usadas para investigações mais aprofundadas e respostas rápidas. Isso ajuda a reduzir o tempo de detecção e resposta a incidentes, mitigando potenciais danos.

2.2 Conceitos de Segurança de Servidores Web e Detecção de Anomalias

2.2.1 Definição de servidores web e seu papel na disponibilização de conteúdo online

Servidores web são sistemas de computadores que hospedam e entregam páginas da web para os usuários através da Internet. Eles desempenham um papel

fundamental ao permitir que os sites e aplicativos online sejam acessados pelos navegadores dos usuários. Os servidores web recebem solicitações de clientes, como navegadores da web, e respondem fornecendo o conteúdo solicitado, que pode ser páginas HTML, imagens, arquivos de vídeo, aplicativos web, entre outros.

2.2.2 Noções fundamentais de segurança de servidores web: autenticação, autorização e criptografia

A segurança de servidores web é crucial para proteger os dados dos usuários, a integridade do conteúdo e prevenir o acesso não autorizado. A autenticação envolve a verificação da identidade do usuário antes de conceder acesso aos recursos do servidor. A autorização, por sua vez, lida com a concessão de permissões apropriadas aos usuários autenticados, controlando quais ações eles podem realizar. A criptografia desempenha um papel fundamental na proteção dos dados em trânsito, garantindo que as informações trocadas entre o cliente e o servidor sejam ilegíveis para terceiros não autorizados.

Segundo a OWASP (Open Web Application Security Project), "A autenticação é o processo de estabelecer confiantemente a identidade de um usuário, geralmente ao confirmar um conjunto de credenciais fornecidas por esse usuário". (OWASP, 2021)

2.2.3 Conceitos de detecção de anomalias em servidores web: identificação de padrões suspeitos e comportamentos não usuais

A detecção de anomalias em servidores web envolve monitorar o tráfego e as atividades para identificar padrões incomuns ou comportamentos que desviem do normal. Isso pode incluir taxas de solicitação anormalmente altas, tipos de solicitações incomuns, acesso a áreas restritas sem autorização, entre outros. A detecção de anomalias se baseia em modelos de comportamento típico e utiliza algoritmos de aprendizado de máquina para identificar desvios significativos.

A SANS Institute descreve a detecção de anomalias como "a identificação de atividades que diferem do comportamento previamente estabelecido e que podem representar uma ameaça à segurança". (SANS, 2010)

2.2.4 Importância da detecção de anomalias na proteção contra ameaças cibernéticas

A detecção de anomalias desempenha um papel crítico na defesa contra ameaças cibernéticas, pois muitos ataques e atividades maliciosas não se encaixam em padrões normais de tráfego. Ao identificar comportamentos anômalos, as organizações podem responder rapidamente a possíveis intrusões, ataques de negação de serviço, vazamentos de dados e outras atividades maliciosas. Isso permite uma abordagem proativa para mitigar as ameaças antes que causem danos significativos.

O relatório Verizon Data Breach Investigations Report afirma: "A detecção de anomalias, embora muitas vezes subestimada, é um componente fundamental da detecção de incidentes. As atividades maliciosas geralmente se manifestam como desvios do padrão". (VERIZON, 2008)

2.3 Fundamentos da Detecção de Anomalias

A detecção de anomalias é um componente vital da segurança cibernética que se concentra na identificação de comportamentos, eventos ou padrões que desviam significativamente do normal. Ela desempenha um papel crucial na proteção de sistemas e redes contra ameaças cibernéticas, pois muitos ataques e atividades maliciosas não seguem padrões previsíveis. A detecção de anomalias visa identificar atividades suspeitas ou inesperadas que possam indicar a presença de ameaças, violações de segurança ou comportamentos anormais. (Chandola, 2009)

2.3.1 Explanação sobre o conceito de detecção de anomalias

O conceito de detecção de anomalias envolve a identificação de desvios significativos em relação ao comportamento normal. Isso pode ser aplicado em diversos contextos, como detecção de fraudes financeiras, monitoramento de sistemas industriais e, especialmente, segurança de TI. Em sistemas de segurança cibernética, a detecção de anomalias implica em monitorar constantemente as atividades e compará-las com um perfil de comportamento normal, a fim de identificar padrões que não se encaixam nesse perfil.

2.3.2 Divisão de elementos em conjuntos "normais" e "anômalos" para alertar sobre indivíduos novos e anômalos

A detecção de anomalias envolve a criação de um modelo que define o comportamento considerado "normal" com base em dados históricos. Qualquer desvio significativo em relação a esse modelo é considerado uma anomalia. Isso pode ser ilustrado usando um exemplo de uma população de transações financeiras. A maioria das transações será rotulada como "normais", mas algumas podem se desviar do padrão e serem identificadas como "anômalas", possivelmente indicando atividades fraudulentas.

2.3.3 Exemplificação por meio de uma população e conjuntos representativos

Considere um sistema de detecção de intrusões em uma rede de computadores. Para estabelecer o comportamento normal, o sistema coleta dados sobre atividades de rede ao longo do tempo. Padrões de tráfego, protocolos usados e tipos de comunicação são analisados para criar um perfil de comportamento normal. Quando novas atividades ocorrem, o sistema compara essas atividades com o perfil existente. Se uma atividade significativamente diferente for detectada, o sistema pode gerar um alerta indicando uma possível anomalia, que pode ser uma tentativa de invasão.

2.3.4 Necessidade de detecção em tempo real para intervenção efetiva

A detecção de anomalias é mais eficaz quando realizada em tempo real. Isso ocorre porque muitas ameaças cibernéticas podem se espalhar e causar danos rapidamente. A detecção em tempo real permite uma resposta imediata, ajudando a interromper atividades maliciosas antes que causem danos substanciais. Por exemplo, a detecção de um padrão anômalo de tráfego em um servidor web pode permitir que as medidas de mitigação sejam tomadas antes que um ataque de negação de serviço seja bem-sucedido. (Gao, 2017)

2.4 Automatização da Detecção de Anomalias: Motivação e Benefícios

A automatização da detecção de anomalias é uma abordagem crítica para lidar com os desafios crescentes apresentados pela complexidade das redes, o aumento exponencial dos volumes de dados e a natureza evolutiva das ameaças cibernéticas.

Ao integrar algoritmos de aprendizado de máquina e técnicas de análise de dados, essa abordagem oferece uma série de benefícios significativos que fortalecem a postura de segurança das organizações.

2.4.1 Tratamento de grandes volumes de dados e a evolução das redes

A automatização da detecção de anomalias tornou-se uma necessidade à medida que as redes e sistemas cibernéticos continuam a crescer em tamanho e complexidade. Com a proliferação de dispositivos conectados à Internet das Coisas (IoT) e o aumento da quantidade de dados gerados, é impraticável depender apenas de monitoramento manual para identificar comportamentos suspeitos. Como mencionado por Han et al., "a automação é crucial para lidar com o crescente volume de dados gerados por sistemas de rede". A detecção manual em ambientes tão dinâmicos e em constante evolução é inviável e ineficaz. (Han, 2020)

2.4.2 Redução de erros humanos e avaliação dos mecanismos utilizados

A automatização elimina a possibilidade de erros humanos decorrentes de fatores como fadiga, falta de atenção ou limitações cognitivas. Além disso, a automação permite a implementação de algoritmos sofisticados e precisos que podem identificar padrões sutis e complexos que os analistas humanos podem perder. Isso é particularmente importante no contexto da detecção de anomalias, onde a identificação precoce e precisa é essencial para mitigar ameaças.

2.4.3 Criação de um sistema autônomo baseado em desvios de políticas predefinidas

A automação da detecção de anomalias permite a criação de sistemas autônomos que operam com base em políticas predefinidas. Os modelos de comportamento normal são construídos a partir de dados históricos e as políticas definem os limites aceitáveis de desvio. Quando o sistema detecta atividades que se desviam dessas políticas, alertas podem ser gerados automaticamente. Como afirmado por Ahad et al., "sistemas automatizados podem tomar decisões mais rápidas e precisas ao comparar eventos observados com políticas predefinidas". (Ahad, 2019)

2.4.4 Detecção de desvio de comportamento normal e ativação de mecanismos preventivos

A automação da detecção de anomalias não apenas identifica atividades suspeitas, mas também pode ser integrada com mecanismos preventivos. Quando um comportamento anômalo é identificado, o sistema pode acionar medidas proativas, como bloqueio de tráfego, isolamento de sistemas ou notificação automática de administradores. Isso permite que as organizações reajam rapidamente a potenciais ameaças, reduzindo o tempo de exposição a ataques.

2.5 Técnicas de Análise de Séries Temporais Aplicadas à Detecção de Anomalias em Tempo de Resposta de Servidores Web

A análise de séries temporais é uma abordagem fundamental para compreender e modelar dados que variam ao longo do tempo. Quando aplicada à detecção de anomalias em servidores web, essa técnica oferece insights valiosos para identificar comportamentos anômalos e variações nos tempos de resposta, auxiliando na manutenção da disponibilidade e desempenho dos sistemas. (Brockwell, 2016)

2.5.1 Introdução à análise de séries temporais

A análise de séries temporais envolve o estudo e a modelagem de dados que são coletados sequencialmente ao longo do tempo. Ela é amplamente utilizada em várias disciplinas, incluindo finanças, ciências sociais e ciência da computação. Na detecção de anomalias em servidores web, a análise de séries temporais permite identificar padrões sazonais, tendências e flutuações temporais nos tempos de resposta, fornecendo uma base para a detecção de comportamentos anormais.

2.5.2 Aplicações da análise de séries temporais na detecção de anomalias em servidores web

A aplicação da análise de séries temporais na detecção de anomalias em servidores web é especialmente útil para monitorar as variações nos tempos de resposta das solicitações. Com uma quantidade crescente de tráfego da web, é fundamental identificar alterações abruptas nos padrões de tempo de resposta que possam indicar atividades maliciosas ou problemas de desempenho. A análise de

séries temporais pode revelar comportamentos que não seriam detectados facilmente por outras abordagens de segurança.

2.5.3 Métodos estatísticos e algoritmos de aprendizado de máquina para análise de séries temporais

A análise de séries temporais na detecção de anomalias em servidores web pode empregar uma variedade de métodos estatísticos e algoritmos de aprendizado de máquina. Métodos clássicos, como modelos ARIMA (Autoregressive Integrated Moving Average) ou decomposição sazonal, podem ser usados para capturar padrões temporais. Além disso, algoritmos de aprendizado de máquina, como redes neurais recorrentes (RNNs) ou algoritmos baseados em árvores, podem ser aplicados para detectar anomalias em padrões complexos.

2.5.4 Casos de uso da análise de séries temporais na detecção de variações de tempo de resposta em servidores web

A análise de séries temporais na detecção de variações de tempo de resposta em servidores web pode ser usada para identificar picos anormais de tráfego que podem indicar um ataque de negação de serviço (DDoS), flutuações sazonais nos tempos de resposta ou mesmo variações inesperadas nos horários de maior atividade. A detecção dessas anomalias permite que as equipes de segurança ou administração de sistemas tomem medidas proativas para garantir a disponibilidade e o desempenho contínuos dos servidores. (Lippi, 2018)

2.6 Estatísticas de Violações de Segurança da Informação e Privacidade: Tendências e Impacto

As ameaças cibernéticas e as violações de segurança da informação se tornaram uma preocupação crescente na era digital. O panorama atual dessas ameaças revela uma constante evolução das táticas dos invasores, tornando essencial o entendimento das estatísticas por trás dos ataques cibernéticos e das violações de privacidade.

2.6.1 Panorama atual das ameaças cibernéticas e violações de segurança da informação

O cenário das ameaças cibernéticas é complexo e dinâmico, com atores maliciosos utilizando uma variedade de métodos para comprometer sistemas, roubar dados e interromper serviços. Ataques como *phishing*, *ransomware*, ataques de negação de serviço (DDoS) e exploração de vulnerabilidades são apenas alguns exemplos das táticas empregadas. Além disso, a natureza sofisticada desses ataques é evidenciada pelo aumento da utilização de técnicas como engenharia social e ataques direcionados.

2.6.2 Estatísticas de ataques cibernéticos e suas implicações para a segurança de servidores web

As estatísticas de ataques cibernéticos destacam a escala dessas ameaças. Relatórios frequentes de ataques bem-sucedidos a grandes empresas e organizações governamentais ressaltam a necessidade de medidas de segurança robustas. Por exemplo, um relatório da Accenture revelou que, em 2020, as violações de segurança aumentaram em 11% e os custos médios das violações aumentaram em 31%. Isso ilustra a urgência de proteger servidores web e sistemas relacionados, que muitas vezes atuam como pontos de entrada para invasões. (Accenture, 2021)

2.6.3 Impacto das violações de privacidade e seus efeitos nas organizações

As violações de privacidade têm repercussões significativas, afetando a confiança dos clientes, a reputação das empresas e, muitas vezes, resultando em multas substanciais por violações regulatórias, como o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia. As consequências financeiras e legais podem ser devastadoras para as organizações. Por exemplo, a Equifax, uma agência de crédito dos EUA, sofreu uma violação em 2017 que afetou 147 milhões de pessoas e resultou em um acordo de mais de 575 milhões de dólares para resolver reclamações e multas. (FTC, 2019)

2.7 Métodos e Técnicas de Prevenção de Incidentes de Segurança em Servidores Web

Garantir a segurança dos servidores web é essencial para proteger a integridade dos dados, a privacidade dos usuários e a disponibilidade dos serviços. Uma abordagem abrangente envolve a implementação de estratégias de proteção, o

uso de técnicas avançadas de monitoramento e a adoção de práticas de manutenção proativas.

2.7.1 Estratégias de proteção de servidores web

As estratégias de proteção de servidores web englobam uma série de medidas para minimizar as vulnerabilidades e proteger contra ameaças. Firewalls, antivírus e sistemas de prevenção de intrusão (IPS) são exemplos de tecnologias que podem ser implantadas para filtrar o tráfego malicioso e identificar comportamentos suspeitos. Os firewalls, por exemplo, atuam como barreiras de entrada, controlando o tráfego de entrada e saída e bloqueando acessos não autorizados. A utilização dessas tecnologias em camadas pode proporcionar uma defesa mais robusta contra ameaças variadas. (Mirkovic, 2017)

2.7.2 Técnicas de monitoramento de tráfego e análise de logs para identificação precoce de ameaças

O monitoramento contínuo do tráfego e a análise dos registros (logs) são técnicas valiosas para identificar atividades anômalas e potencialmente maliciosas. O monitoramento do tráfego permite a detecção de padrões de tráfego incomuns ou picos de atividade que podem indicar um ataque em andamento. A análise de logs fornece informações detalhadas sobre as atividades no servidor, ajudando a identificar tentativas de invasão, acessos não autorizados e outras atividades suspeitas. Essas técnicas são cruciais para a detecção precoce de ameaças e a resposta eficaz a incidentes de segurança.

2.7.3 Importância da atualização de software e patches de segurança para prevenir vulnerabilidades

Manter o software e os sistemas atualizados é uma prática essencial para prevenir vulnerabilidades conhecidas. Muitos ataques exploram falhas de segurança em software desatualizado. A atualização regular do sistema operacional, dos aplicativos e das bibliotecas utilizadas no servidor web é fundamental para mitigar riscos. Além disso, a aplicação de patches de segurança disponibilizados pelos fornecedores é crucial para fechar potenciais brechas de segurança. (Cert.br, s.d.)

2.8 Normas e Regulamentações Nacionais e Internacionais Relacionadas à Segurança de Servidores Web e Detecção de Anomalias

Em um mundo cada vez mais digital e interconectado, a proteção da segurança da informação e da privacidade tornou-se uma preocupação central. Normas e regulamentações nacionais e internacionais têm sido desenvolvidas para orientar as organizações na implementação de medidas de segurança eficazes, incluindo a proteção de servidores web e a detecção de anomalias.

2.8.1 Principais normas de segurança da informação e privacidade

Normas como a ISO 27001 fornecem diretrizes abrangentes para o estabelecimento, implementação, operação, monitoramento, revisão, manutenção e melhoria de um sistema de gestão de segurança da informação. Elas definem controles e boas práticas para garantir a confidencialidade, integridade e disponibilidade das informações. A Regulação Geral de Proteção de Dados (GDPR) da União Europeia estabelece regras rigorosas para a proteção de dados pessoais. Além disso, regulamentações específicas, como a Lei de Portabilidade e Responsabilidade de Seguro de Saúde (HIPAA) nos Estados Unidos, focam em proteger informações médicas e de saúde.

2.8.1.1 ISO 27001: Sistema de Gestão de Segurança da Informação (SGSI)

A ISO 27001 é uma norma internacional que fornece um quadro abrangente para estabelecer, implementar, operar, monitorar, revisar, manter e melhorar um Sistema de Gestão de Segurança da Informação (SGSI). Ela se concentra na gestão proativa da segurança da informação e engloba a aplicação de controles e práticas para garantir a confidencialidade, integridade e disponibilidade das informações. A norma é amplamente adotada por organizações de diversos setores para mitigar riscos de segurança da informação e proteger ativos valiosos. (ISO/IEC, 2021)

2.8.1.2 Regulação Geral de Proteção de Dados (GDPR)

A GDPR é uma regulamentação da União Europeia que estabelece regras rigorosas para a proteção de dados pessoais dos cidadãos europeus. Ela abrange a coleta, processamento e armazenamento de dados pessoais, exigindo que as organizações adotem medidas de segurança robustas e transparentes. A GDPR

também garante aos indivíduos maior controle sobre seus dados e introduz penalidades substanciais para violações de privacidade. (GDPR, 2018)

2.8.1.3 Lei de Portabilidade e Responsabilidade de Seguro de Saúde (HIPAA)

Nos Estados Unidos, a Lei de Portabilidade e Responsabilidade de Seguro de Saúde (HIPAA) estabelece regulamentações específicas para a proteção de informações médicas e de saúde. A HIPAA exige que as organizações de saúde implementem salvaguardas técnicas, administrativas e físicas para proteger informações confidenciais de pacientes. Essa regulamentação tem como objetivo garantir a privacidade dos dados médicos e promover a segurança dos sistemas de informação na área da saúde. (HIPAA, 2021)

2.8.1.4 Lei Geral de Proteção de Dados (LGPD)

O Brasil também possui regulamentações e normas específicas que desempenham um papel importante na segurança da informação e na privacidade dos dados. É o caso da Lei Geral de Proteção de Dados (LGPD) é a regulamentação brasileira que estabelece regras para a coleta, processamento e armazenamento de dados pessoais. Inspirada na GDPR da União Europeia, a LGPD tem como objetivo proteger a privacidade dos cidadãos brasileiros, dando-lhes maior controle sobre seus dados e impondo responsabilidades às organizações que lidam com informações pessoais. A LGPD também exige que as empresas adotem medidas de segurança para proteger os dados que coletam e processam. (Brasil, LEI Nº 13.709, DE 14 DE AGOSTO DE 2018, 2018)

2.8.1.5 Marco Civil da Internet

O Marco Civil da Internet é uma lei brasileira que estabelece os princípios, direitos e deveres para o uso da internet no Brasil. Ele visa garantir a liberdade de expressão, a privacidade dos usuários e a neutralidade da rede, além de fornecer orientações para a responsabilidade de provedores de serviços e a proteção de dados. O Marco Civil também inclui diretrizes para a cooperação entre governos e empresas em questões de segurança cibernética. (Brasil, LEI Nº 12.965, DE 23 DE ABRIL DE 2014, 2014)

2.8.1.6 Agência Nacional de Segurança Cibernética (ANSC)

A Agência Nacional de Segurança Cibernética (ANSC) é uma instituição brasileira criada para promover a segurança cibernética no país. A ANSC atua na coordenação de ações de prevenção, detecção, resposta e mitigação de incidentes cibernéticos. Ela desempenha um papel fundamental na coordenação de esforços entre entidades públicas e privadas para proteger a infraestrutura crítica e promover a conscientização sobre ameaças cibernéticas. (ANSC, 2023)

2.8.1.7 Regulamentações específicas para setores

Além das regulamentações mencionadas acima, existem regulamentações específicas para setores como serviços financeiros (Banco Central do Brasil), telecomunicações (Agência Nacional de Telecomunicações - ANATEL) e saúde (Agência Nacional de Saúde Suplementar - ANS).

Essas regulamentações têm como objetivo garantir a segurança e a proteção dos dados em setores sensíveis, estabelecendo diretrizes específicas para as organizações que operam nesses segmentos.

Em conjunto, essas regulamentações e normas brasileiras refletem o compromisso do país em proteger a segurança da informação e a privacidade dos dados. Elas definem padrões, diretrizes e responsabilidades para organizações e indivíduos, promovendo um ambiente digital mais seguro e confiável.

2.8.2 Regulamentações específicas para servidores web

Garantir a segurança de servidores web é essencial para proteger a integridade das aplicações, dados e informações sensíveis armazenadas e transmitidas por meio desses sistemas. Diversas entidades e regulamentações específicas têm sido desenvolvidas para fornecer diretrizes detalhadas e melhores práticas para proteger servidores web contra ameaças cibernéticas.

O Instituto Nacional de Padrões e Tecnologia (NIST) e a Open Web Application Security Project (OWASP) são exemplos de entidades que fornecem diretrizes específicas para a segurança de servidores web. O NIST publica documentos como o "Guia de Segurança para Servidores Web" que detalham práticas recomendadas para proteger servidores web contra ameaças cibernéticas.

O OWASP, por sua vez, mantém projetos e recomendações centradas na segurança de aplicações web, incluindo servidores web. O Padrão de Segurança de Dados para a Indústria de Cartões de Pagamento (PCI DSS) é uma regulamentação voltada para a proteção de dados de pagamento e aplica-se a servidores web que lidam com transações financeiras.

2.8.2.1 Instituto Nacional de Padrões e Tecnologia (NIST)

O NIST é uma agência do governo dos Estados Unidos que desempenha um papel fundamental no estabelecimento de padrões e diretrizes de segurança cibernética. O NIST publica documentos como o "Guia de Segurança para Servidores Web", que fornece práticas recomendadas para proteger servidores web contra uma ampla gama de ameaças. Esse guia aborda aspectos como configuração segura de servidores, gerenciamento de identidade e acesso, proteção contra injeções de código, e outros tópicos críticos para a segurança de aplicações web. (NIST, 2023)

2.8.2.2 Open Web Application Security Project (OWASP)

O OWASP é uma comunidade global que se concentra em melhorar a segurança de software. Eles mantêm projetos, guias e recomendações específicas para a segurança de aplicações web, incluindo servidores web. O Projeto OWASP Top Ten lista as dez vulnerabilidades mais críticas que afetam aplicações web, como injeção de SQL, cross-site scripting (XSS) e autenticação inadequada. Essas diretrizes auxiliam os desenvolvedores e administradores na identificação e mitigação dessas vulnerabilidades em servidores web. (OWASP, s.d.)

2.8.2.3 Padrão de Segurança de Dados para a Indústria de Cartões de Pagamento (PCI DSS)

O PCI DSS é um conjunto de regulamentações voltado para a segurança de dados de pagamento. Ele se aplica a organizações que processam, armazenam ou transmitem dados de cartões de pagamento. Uma parte significativa das transações financeiras ocorre online, e servidores web desempenham um papel central nessas operações. O PCI DSS define diretrizes rigorosas para proteger informações de pagamento, exigindo a implementação de medidas de segurança, auditorias e testes regulares para garantir a conformidade e a proteção dos dados financeiros. (PSI, 2023)

2.8.3 Impacto das regulamentações na segurança e detecção de anomalias em servidores web

Conforme afirmado por (FERREIRA, 2017), as regulamentações desempenham um papel fundamental na promoção de práticas de segurança cibernética eficazes e na detecção de anomalias em servidores web.

Conforme destacado por essas entidades, (Security Configuration Checklists Program for IT Products: Guidelines for Web Servers. Fonte: NIST: , 2021)essas diretrizes são fundamentais para estabelecer medidas de segurança eficazes. Elas estabelecem padrões mínimos e diretrizes que as organizações devem seguir para garantir a proteção de dados, informações sensíveis e a disponibilidade de serviços. O impacto dessas regulamentações é significativo, abordando aspectos que vão desde a implementação de medidas de segurança até a detecção precoce de atividades suspeitas.

2.8.3.1 Definição de Padrões Mínimos de Segurança

As regulamentações estabelecem padrões mínimos de segurança que as organizações devem cumprir para proteger seus ativos digitais e informações. Isso inclui a implementação de medidas de proteção de servidores web, como firewalls, sistemas de detecção de intrusão (IDS), criptografia de dados e autenticação robusta. Essas medidas ajudam a prevenir vulnerabilidades e ataques direcionados aos servidores web, protegendo contra perdas financeiras e de reputação.

2.8.3.2 Implementação de Detecção de Anomalias

Muitas regulamentações incentivam ou exigem a implementação de sistemas de detecção de anomalias em servidores web. Isso envolve a monitorização constante do tráfego, a análise de logs e a identificação de comportamentos suspeitos. A detecção de anomalias pode indicar tentativas de intrusão, exploração de vulnerabilidades ou atividades maliciosas, permitindo uma resposta rápida antes que danos significativos ocorram.

2.8.3.3 Redução do Risco de Multas e Penalidades

O não cumprimento das regulamentações pode resultar em multas substanciais e outras penalidades. Ao implementar medidas de segurança e detecção de

anomalias conforme exigido pelas regulamentações, as organizações reduzem o risco de violações e as consequências financeiras associadas. O cumprimento das regulamentações também demonstra compromisso com a proteção dos dados e a conformidade legal.

2.8.3.4 Melhoria da Conscientização e Cultura de Segurança

As regulamentações também promovem uma cultura de segurança cibernética mais ampla dentro das organizações. A necessidade de cumprir regulamentações leva as empresas a investirem em treinamento, conscientização e educação dos funcionários sobre boas práticas de segurança. Isso contribui para uma compreensão mais sólida das ameaças cibernéticas e ações proativas para preveni-las.

2.8.3.5 Promoção da Confiança dos Clientes e Parceiros

O cumprimento das regulamentações de segurança cibernética e detecção de anomalias não apenas protege os ativos internos, mas também aumenta a confiança dos clientes, parceiros comerciais e outras partes interessadas. Mostra que a organização está comprometida em proteger as informações confidenciais e a privacidade dos usuários, fortalecendo a reputação e a credibilidade.

3 METODOLOGIA

Esta metodologia foi inspirada em uma iniciativa do projeto colaborativo entre o Communications Security Establishment (CSE) e o Canadian Institute for Cybersecurity (CIC), que usa a noção de perfis para gerar um conjunto de dados de segurança cibernética de maneira sistemática, chamado de CSE-CIC-IDS2018. (Registry of Open Data on AWS, 2023)

Ele inclui uma descrição detalhada de invasões junto com modelos de distribuição abstratos para aplicativos, protocolos ou entidades de rede de nível inferior. O conjunto de dados inclui sete cenários de ataque diferentes, ou seja, força bruta, Heartbleed, Botnet, DoS, DDoS, ataques da Web e infiltração da rede de dentro.

A infra-estrutura de ataque inclui 50 máquinas e a organização vítima tem 5 departamentos, incluindo 420 PCs e 30 servidores. Este conjunto de dados inclui o tráfego de rede e os arquivos de log de cada máquina do lado da vítima, juntamente com 80 recursos de tráfego de rede extraídos do tráfego capturado usando o CICFlowMeter-V3.

Para obter mais informações sobre a criação desse conjunto de dados, consulte este artigo de pesquisadores do CIC e da University of New Brunswick (UNB)

Neste sentido, inspirado pelo dataset do CSE, foi preparado um arquivo de texto “tempo-de-resposta-1.txt” para facilitar a geração das análises, limitado a um número menor de ocorrências. Este arquivo contém:

- 135.149 valores de tempo de resposta (em milissegundos) de um servidor web
- Medições feitas em intervalos de 5 segundos
- Aproximadamente 8 dias de medições

Considerando também que, o tempo de resposta tende a ser influenciado pela quantidade de acessos simultâneos, partiu-se da hipótese de que um ataque de negação de serviço tende a elevar o tempo de resposta para usuários legítimos.

- Menor valor medido: 3,025ms
- Maior valor medido: 2069,171ms
- Valor médio calculado: 9,83ms

Partindo da premissa de que o tempo de resposta é afetado pelo número de acessos simultâneos, levantamos a hipótese de que um ataque de negação de serviço pode aumentar o tempo de resposta para os usuários legítimos.

O método proposto consiste em categorizar a população de tempos de resposta em "normais" e "anômalos," alertando sobre a detecção de novos indivíduos anômalos. Dada uma população P,

$$P = \{80,70,90,80,2000,80,80,5\}$$

tem-se o objetivo em dividir os elementos de P em um conjunto N de elementos “normais”

$$N = \{80,70,90,80,80,80\}$$

e outro O, de indivíduos “anômalos”, alertando sempre que um novo indivíduo anômalo surgir.

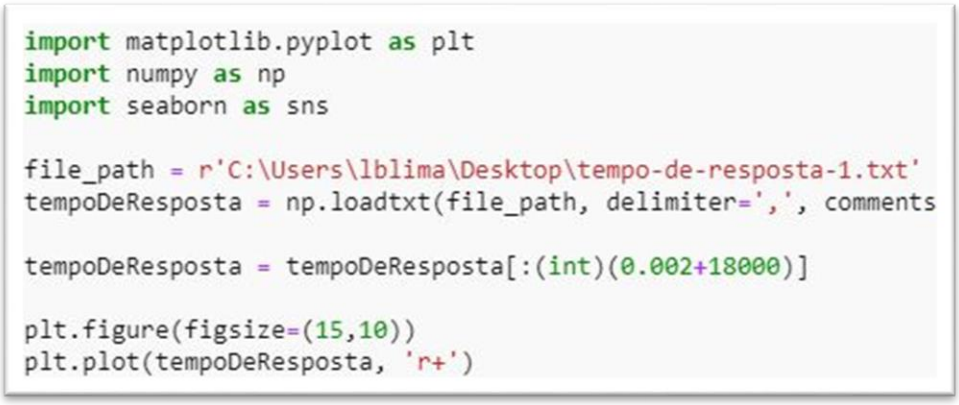
$$O = \{5,2000\}$$

Para facilitar o processo, estabelecemos um limite de amostragem de 18.000 para a análise.

3.1 Parte 1: Carregamento dos Dados

Nesta etapa, foi realizado o carregamento dos dados. Abaixo estão as instruções para carregamento do arquivo "tempo-de-resposta-1.txt", esse arquivo contém informações sobre o tempo de resposta de requisições a um servidor web. É importante ressaltar que esses valores podem ter sido obtidos por meio de uma rotina periódica, que tem como objetivo avaliar a "saúde" do servidor web.

Figura 12: Importação das bibliotecas e carregamento dos dados



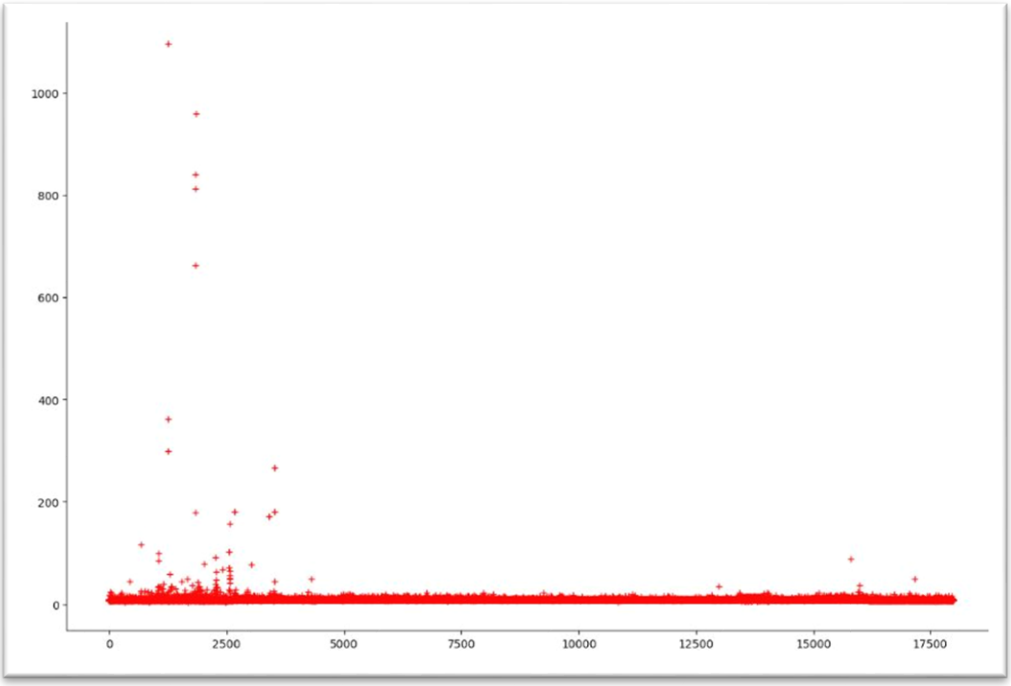
```
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns

file_path = r'C:\Users\lblima\Desktop\tempo-de-resposta-1.txt'
tempoDeResposta = np.loadtxt(file_path, delimiter=',', comments

tempoDeResposta = tempoDeResposta[:(int)(0.002+18000)]

plt.figure(figsize=(15,10))
plt.plot(tempoDeResposta, 'r+')
```

Fonte: (Bruscagini, 2023)

Aqui a variável tempoDeResposta = tempoDeResposta[:(int)(0.002+18000)] limita o tamanho do conjunto de dados para os primeiros 18000 elementos. Esse processamento é realizado utilizando o slicing do NumPy, onde tempoDeResposta[:(int)(0.002+18000)] seleciona os primeiros 18000 elementos da matriz.

A seguir, o código cria uma figura de tamanho 15x10 polegadas com plt.figure(figsize=(15,10)) e, em seguida, plota o conjunto de dados no gráfico utilizando plt.plot(tempoDeResposta, 'r+'). Nessa visualização, os dados de tempo de resposta são plotados com pontos vermelhos ('r+') em relação ao índice dos elementos da matriz, formando um gráfico de linha.

Gráfico 4: Elementos da matriz tempo de resposta X Instantes de tempo



Fonte: (Bruscagini, 2023)

A escala no eixo vertical é determinada automaticamente pelo matplotlib com base nos valores mínimos e máximos dos dados de tempo de resposta. O matplotlib faz o ajuste automático da escala para que todos os pontos do gráfico sejam visíveis dentro dos limites do eixo. Isso significa que, se os valores de tempo de resposta variarem significativamente, a escala vertical se ajustará para acomodar todos os pontos do gráfico, garantindo que nenhum ponto seja cortado ou omitido.

Lembrando que, no eixo horizontal, os valores representam os índices dos elementos da matriz tempoDeResposta. Cada ponto no gráfico está associado a um índice específico da matriz, o que permite visualizar a sequência dos valores de tempo de resposta.

E, no eixo vertical, os valores representam os dados de tempo de resposta contidos na matriz tempoDeResposta. Esses valores são plotados em relação aos índices no eixo horizontal, ou seja, cada ponto no gráfico corresponde a um valor específico do tempo de resposta.

3.2 Parte 2: Cálculo da Média e Identificação do Comportamento do Servidor Web

O cálculo da média é uma ferramenta crucial para compreender o comportamento do servidor web à medida que os valores são medidos pelo sistema de monitoramento de sua "saúde". O objetivo central é obter uma estimativa do padrão médio do servidor.

Uma vez que diferentes valores são capturados em intervalos regulares, a média é continuamente recalculada para refletir as variações observadas. Ao longo do tempo, essa média tende a se estabilizar, fornecendo uma representação consolidada do sistema.

No entanto, é importante considerar a adaptabilidade da média em diferentes cenários. Em algumas situações, pode ser vantajoso focar em um intervalo de tempo mais curto, permitindo uma análise mais imediata. Isso permite uma visão mais precisa do comportamento atual do sistema, mesmo que signifique "esquecer" temporariamente as medições mais antigas.

Portanto, a abordagem de calcular a média e ajustar a janela de análise de acordo com as necessidades oferece uma perspectiva abrangente e dinâmica da "saúde" do servidor web, equilibrando a compreensão de tendências de longo prazo com a identificação de mudanças recentes.

A seguir podemos verificar a matriz medias é inicializada como uma matriz vazia (np.empty(0)). Essa matriz será utilizada para armazenar as médias móveis cumulativas calculadas ao longo do loop.

Figura 13: Cálculo da média móvel cumulativa

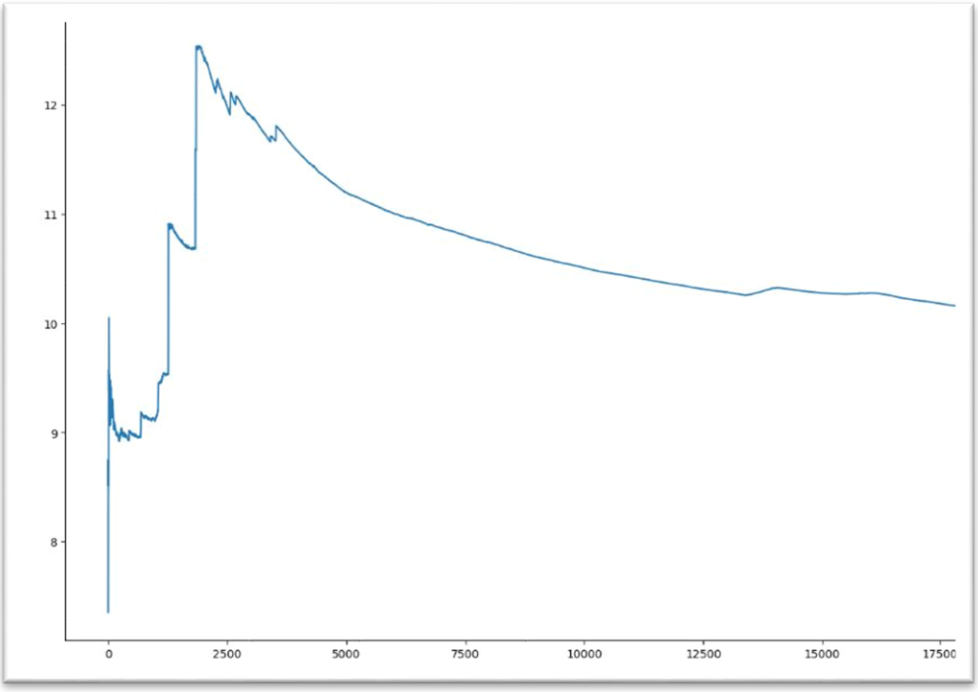
```
medias = np.empty(0)
for i in range(len(tempoDeResposta)):
  medias=np.append(medias, np.mean(tempoDeResposta[0:i+1]))

plt.figure(figsize=(15,10))
plt.plot(medias)
```

Fonte: (Bruscagini, 2023)

O código utiliza um loop for para percorrer cada elemento da matriz tempoDeResposta. O índice i representa a posição atual no loop, indo de 0 até o tamanho total da matriz menos 1 (len(tempoDeResposta) - 1). Dentro do loop, o código utiliza a função np.mean() do NumPy para calcular a média dos elementos de tempoDeResposta até o índice atual (i+1). Em outras palavras, ele calcula a média dos primeiros i+1 elementos de tempoDeResposta, incluindo o elemento atual em cada iteração. A média é armazenada na matriz medias usando a função np.append() do NumPy.

Após calcular todas as médias móveis cumulativas, o código cria uma figura de tamanho 15x10 polegadas com plt.figure(figsize=(15,10)). Em seguida, ele plota as médias móveis cumulativas na forma de um gráfico de linha utilizando plt.plot(medias). O gráfico mostra como as médias variam ao longo do tempo e proporciona uma visão mais suavizada das tendências nos dados originais de tempo de resposta.

Gráfico 5: Representação da média com 41806 anomalias (30.93% do total).



Fonte: (Bruscagini, 2023)

3.3 Parte 3: Explorando o cálculo da média móvel

Dando continuidade à análise, agora abordaremos o cálculo da média de uma forma diferente, considerando o contexto do sistema de monitoramento da "saúde" do

servidor web. Neste cenário, enfocaremos a média móvel, que permite uma abordagem mais sensível às variações recentes, evitando que a média seja mascarada por valores passados.

3.3.1 Cálculo da média móvel dentro de uma janela de tempo específica

A ideia é estabelecer um cálculo de média semelhante, mas desta vez aplicando-o dentro de uma janela de tempo específica. Isso proporciona uma visão mais detalhada das flutuações recentes no comportamento do servidor. A lógica subjacente é evitar que a média seja suavizada demais, o que poderia ocultar pontos atípicos que merecem atenção.

3.3.2 Ponderação do passado para ajuste da média

Uma abordagem alternativa surge ao considerar uma média móvel que não descarta completamente os valores passados. Nessa abordagem, o passado é ponderado, de forma que contribua para a formação da média atual. Isso permite um ajuste mais preciso do comportamento "normal" em relação às condições recentes. Importante notar que, conforme o intervalo de tempo se expande, picos e variações notáveis podem se perder, prejudicando a capacidade de detecção de mudanças significativas.

3.3.3 Manutenção da visibilidade de picos e variações

Uma consideração crucial é manter a visibilidade dos picos e das variações acentuadas. Embora a suavização da média possa ser benéfica em alguns casos, é de interesse preservar a detecção desses picos, que podem indicar situações anômalas ou atividades atípicas. A capacidade de identificar esses eventos excepcionais é essencial para a detecção precoce de problemas e ameaças.

Explorar diferentes abordagens para o cálculo da média móvel dentro do contexto de monitoramento da "saúde" do servidor web permite uma análise mais abrangente e sensível ao comportamento do sistema. A seleção entre esquecer mais do passado ou ponderar seu efeito depende da ênfase na detecção de variações recentes versus a manutenção da visibilidade de picos. A escolha adequada pode aprimorar significativamente a capacidade de identificar comportamentos anômalos e

contribuir para a tomada de decisões informadas em relação à segurança e desempenho do servidor.

Figura 14: Gerador de múltiplos gráficos de médias móveis

```
numGraficos = 10
intervalo = 100

fig, axs = plt.subplots(numGraficos, figsize=(15,10))

for j in range(numGraficos-1):
  mediasMoveis = np.empty(0)
  for i in range(0,len(tempoDeResposta),(j+1)*intervalo):
    mediasMoveis = np.append(mediasMoveis, np.mean(tempoDeResposta[i:i+(j+1)*intervalo]))
  axs[j].plot(mediasMoveis)

# O gráfico anterior das médias sem esquecer nada para comparar com os anteriores
axs[j+1].plot(medias)
```

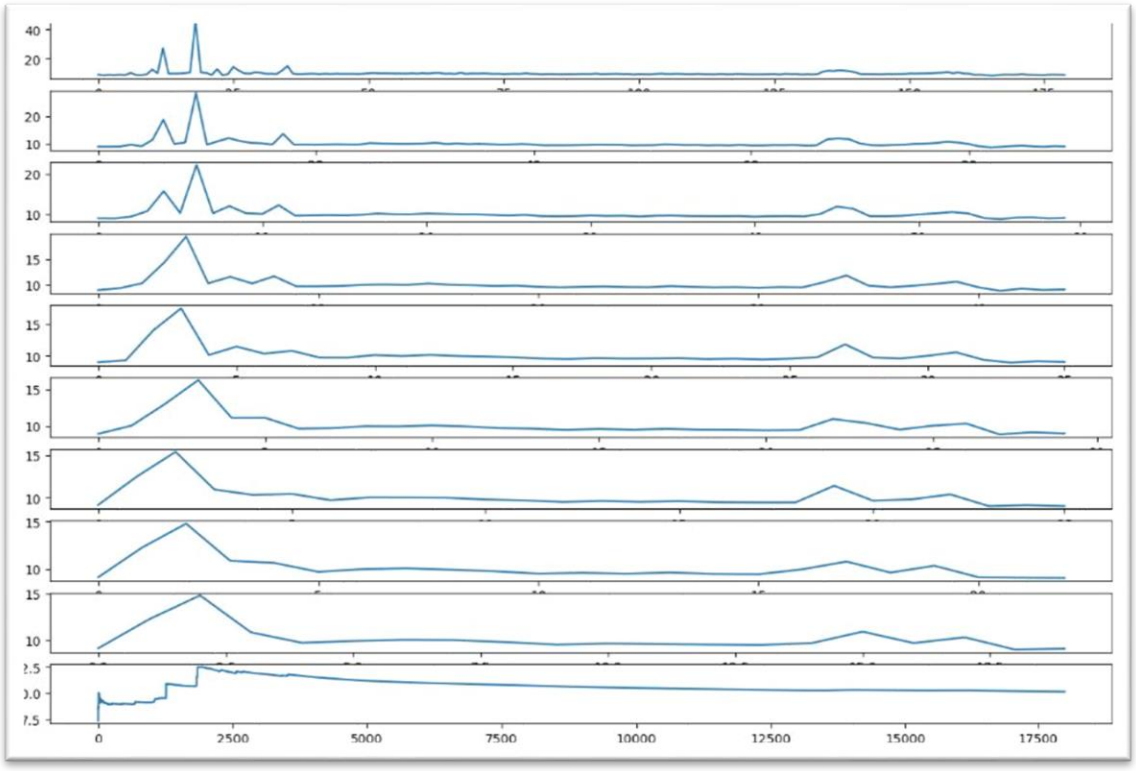
Fonte: (Bruscagini, 2023)

O código entra em um loop for j in range(numGraficos-1), que itera por numGraficos-1 vezes. Para cada iteração, ele calcula as médias móveis usando um determinado intervalo de amostragem.

- mediasMoveis = np.empty(0): Inicializa a matriz mediasMoveis como uma matriz vazia para armazenar as médias móveis calculadas.
- for i in range(0,len(tempoDeResposta),(j+1)*intervalo): Neste loop interno, ele itera sobre o conjunto de dados tempoDeResposta usando um intervalo de amostragem baseado no valor de j. O intervalo de amostragem aumenta a cada iteração em 1, garantindo que as amostras sejam tomadas de forma mais espaçada para calcular as médias móveis em diferentes intervalos.
- mediasMoveis = np.append(mediasMoveis, np.mean(tempoDeResposta[i:i+(j+1)*intervalo])): Calcula a média dos dados de tempoDeResposta dentro do intervalo atual e adiciona essa média à matriz mediasMoveis.
- axs[j].plot(mediasMoveis): Plota o gráfico da média móvel atual no subplot correspondente ao índice j.

Após o loop, o código plota o gráfico das médias sem esquecer nada, provavelmente obtido anteriormente com outro código (não fornecido aqui). O gráfico é plotado no subplot axs[j+1].

Gráfico 6: Médias de 500 em 500, 1000 em 1000, 1500 em 1500.



Fonte: (Bruscagini, 2023)

3.4 Parte 4: Explorando a identificação automatizada de anomalias

No intuito de automatizar a detecção de picos e variações expressivas, os gráficos iniciais das médias móveis se destacam por sua capacidade de realçar esses eventos. No entanto, uma abordagem automatizada é desejada para alcançar essa descoberta, permitindo que a máquina identifique padrões de forma eficiente e precisa.

3.4.1 Busca por padrões automatizados

O uso de gráficos tem seu valor, mas nossa busca agora se concentra na automação do processo. Nosso objetivo é capacitar uma máquina para reconhecer anomalias sem intervenção humana. Para isso, exploraremos a identificação de valores que se desviam de uma faixa definida em torno dos valores mais frequentemente observados.

3.4.2 O papel dos boxplots

Uma abordagem visual interessante para essa tarefa é a utilização de um gráfico conhecido como boxplot. Esse gráfico proporciona uma representação gráfica das distribuições dos dados, permitindo a identificação clara de valores discrepantes. Vamos explorar a aplicação do boxplot nos dados originais para uma compreensão visual da distribuição.

3.4.3 Foco em um intervalo específico

Dado que a visualização de todos os dados em um boxplot completo pode ser complexa, optaremos por uma análise mais focalizada. Geraremos um gráfico boxplot com base em dados específicos, limitados às instâncias entre 4000 e 4100. Isso permitirá uma observação mais nítida dos valores discrepantes e facilitará a identificação de anomalias.

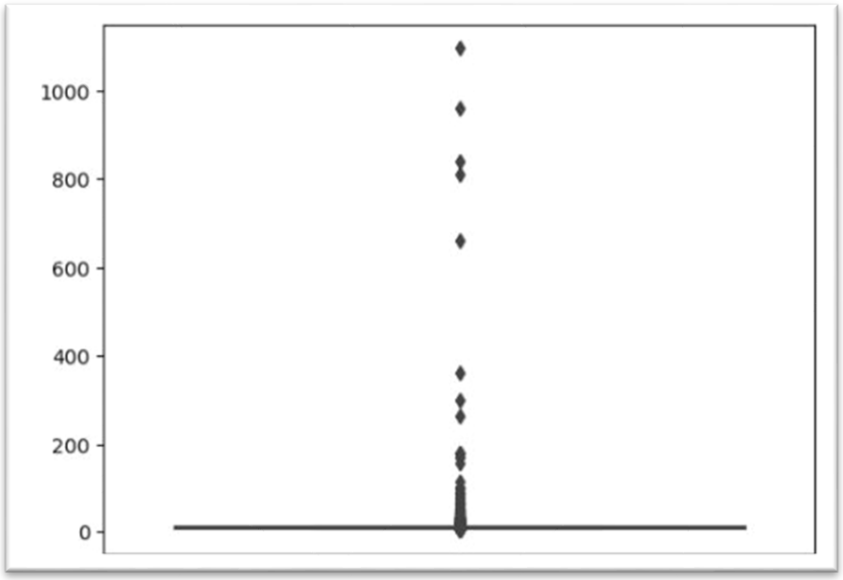
3.4.4 Identificação de anomalias (Outliers)

No contexto da segurança, os pontos que se encontram fora do intervalo delimitado pelas barras do boxplot são considerados como anomalias. Embora sejam frequentemente chamados de outliers, nosso foco na segurança os designará como anomalias. A identificação automatizada desses pontos permitirá a detecção eficaz de comportamentos atípicos, contribuindo para a proteção do servidor web.

Explorando uma abordagem automatizada para a identificação de anomalias, utilizamos gráficos de médias móveis como ponto de partida. Agora, direcionamos nossa atenção para o uso de boxplots para representar visualmente as anomalias nos dados originais. Com a automação desempenhando um papel fundamental, buscamos fortalecer a capacidade do sistema de monitoramento de detectar eventos incomuns e proteger proativamente o servidor web contra ameaças e comportamentos não convencionais.

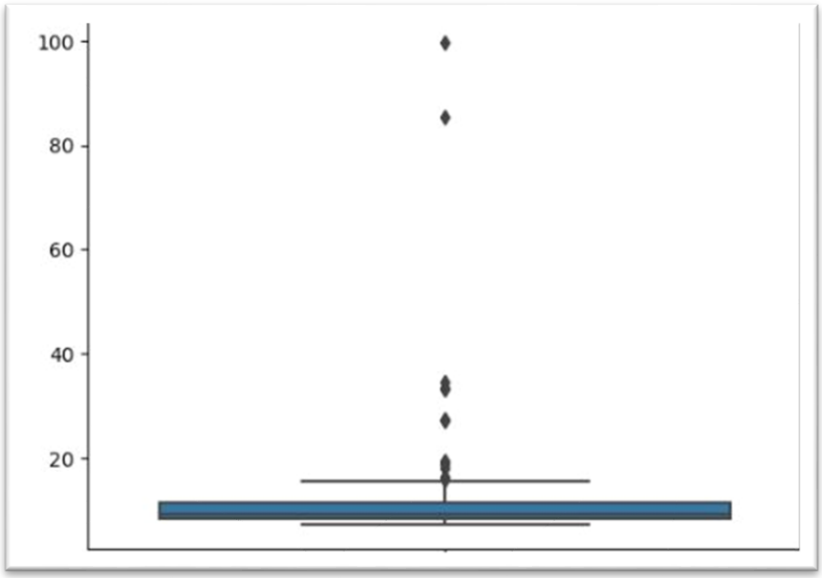
O gráfico de caixa gerado por esse código mostrará a distribuição dos dados de tempoDeResposta em relação à mediana, quartis (25º e 75º percentis) e possíveis outliers (pontos fora dos limites do retângulo). Isso permite uma rápida visualização de como os dados estão dispersos e quaisquer valores atípicos que possam existir.

Gráfico 7: Boxplot de tudo.



Fonte: (Bruscagini, 2023)

Gráfico 8: Boxplot parcial.



Fonte: (Bruscagini, 2023)

Aqui, o gráfico de caixa gerado por esse código mostrará a distribuição dos dados contidos na amostra selecionada (tempoDeResposta[1000:1100]) em relação à mediana, quartis (25º e 75º percentis) e possíveis outliers (pontos fora dos limites do retângulo). Isso permite uma visualização mais detalhada e específica dos dados

contidos nessa parte específica da matriz tempoDeResposta. Neste caso, o parâmetro y indica que os dados a serem usados para criar o gráfico de caixa são os elementos da matriz tempoDeResposta entre os índices 1000 e 1100 (inclusive). Ou seja, a função está selecionando uma amostra específica dos dados, consistindo de 101 elementos (do 1000º ao 1100º elemento, incluindo-os).

3.5 Parte 5: Automatização da detecção e alerta de anomalias

Retornando ao nosso propósito central de automação, agora iremos concentrar nossos esforços em detectar e alertar automaticamente o administrador quando valores fora do intervalo forem identificados no gráfico do boxplot. Nosso objetivo vai além da mera visualização gráfica, buscamos ação e notificação eficazes.

3.5.1 Identificação dos limites no boxplot

O gráfico do boxplot revela os valores que se encontram fora dos limites estabelecidos. No entanto, nossa intenção é identificar esses valores de forma automatizada e comunicar prontamente ao administrador. Para atingir esse objetivo, precisamos compreender como esses limites são definidos. Na parte superior, os valores fora do intervalo do boxplot são delimitados por Q3 + 1,5*IQR.

3.5.2 Quartis e intervalo interquartil (IQR)

O terceiro quartil, Q3, corresponde ao valor até o qual 75% dos valores medidos se agrupam. O primeiro quartil, Q1, representa o valor até o qual 25% dos valores se concentram. O IQR é calculado como Q3 - Q1. Essas métricas são cruciais para a determinação dos limites superiores no boxplot, que indicam a presença de valores discrepantes.

3.5.3 Atualização dos limites

Estabelecemos uma abordagem dinâmica para a definição dos limites no boxplot. Inicialmente, utilizamos os primeiros 2000 valores para calcular o primeiro boxplot. A partir desse ponto, a cada novo valor de tempo de resposta medido, recalculamos os limites dos novos boxplots. Isso permite que os limites se ajustem às variações nos dados ao longo do tempo.

3.5.4 Visualização dos outliers no gráfico original

Uma vez identificados os valores dos outliers, exibimos esses pontos no gráfico original dos tempos de resposta. Essa representação gráfica oferece uma perspectiva visual imediata da precisão do algoritmo e do impacto das detecções automáticas.

Nossa abordagem procura automatizar a detecção de anomalias e proporcionar alertas acionáveis ao administrador. Ao compreender os limites dos boxplots e as métricas de quartis, podemos identificar valores discrepantes e agir de maneira proativa. A atualização constante dos limites e a visualização gráfica dos resultados fortalecem a eficácia do algoritmo de detecção automatizada, permitindo uma abordagem confiável e eficiente na identificação de comportamentos anômalos no servidor web.

Figura 15: Implementação do método para detecção de anomalias

```
numeroDeAnomalias = 0
instantes = []

for i in range(199, len(tempoDeResposta)):
  Q1 = np.percentile(tempoDeResposta[:i+1], 25, interpolation = 'midpoint')
  Q3 = np.percentile(tempoDeResposta[:i+1], 75, interpolation = 'midpoint')
  IQR = Q3 - Q1
  limiteSuperior = Q3 + 1.5*IQR

  if tempoDeResposta[i] > limiteSuperior:
    numeroDeAnomalias+=1
    instantes.append(i)
print(numeroDeAnomalias, "anomalias detectadas (%.2f%% dos valores)" %(100*numeroDeAnomalias/len(tempoDeResposta)))
print("Instantes com anomalias: ", instantes)
```

Fonte: (Bruscagini, 2023)

Este código implementa um método para detectar anomalias nos dados contidos na matriz tempoDeResposta utilizando o conceito de intervalo interquartil (IQR) e a regra do "1.5 vezes o IQR" para identificar valores extremos. Abaixo uma descrição mais detalhada:

numeroDeAnomalias = 0: Inicializa a variável numeroDeAnomalias como 0. Essa variável será usada para contar o número de anomalias encontradas nos dados.

instantes = []: Inicializa a lista instantes, que será usada para armazenar os índices dos instantes em que as anomalias foram detectadas.

3.5.5 Loop para detecção de anomalias:

O código entra em um loop for i in range(199, len(tempoDeResposta)), que itera da posição 199 até o final da matriz tempoDeResposta. A posição 199 é usada como início para garantir que haja dados suficientes para calcular os quartis e o IQR.

- Cálculo dos quartis e IQR:

Para cada iteração do loop, o código calcula o primeiro quartil (Q1) e o terceiro quartil (Q3) dos dados contidos na matriz tempoDeResposta até o índice atual i+1. O IQR é calculado como a diferença entre Q3 e Q1.

- Definição do limite superior:

O limite superior é calculado como limiteSuperior = Q3 + 1.5*IQR. Esse limite é usado para identificar valores considerados extremos, ou seja, valores que estão acima desse limite podem ser considerados anomalias.

- Detecção de anomalias:

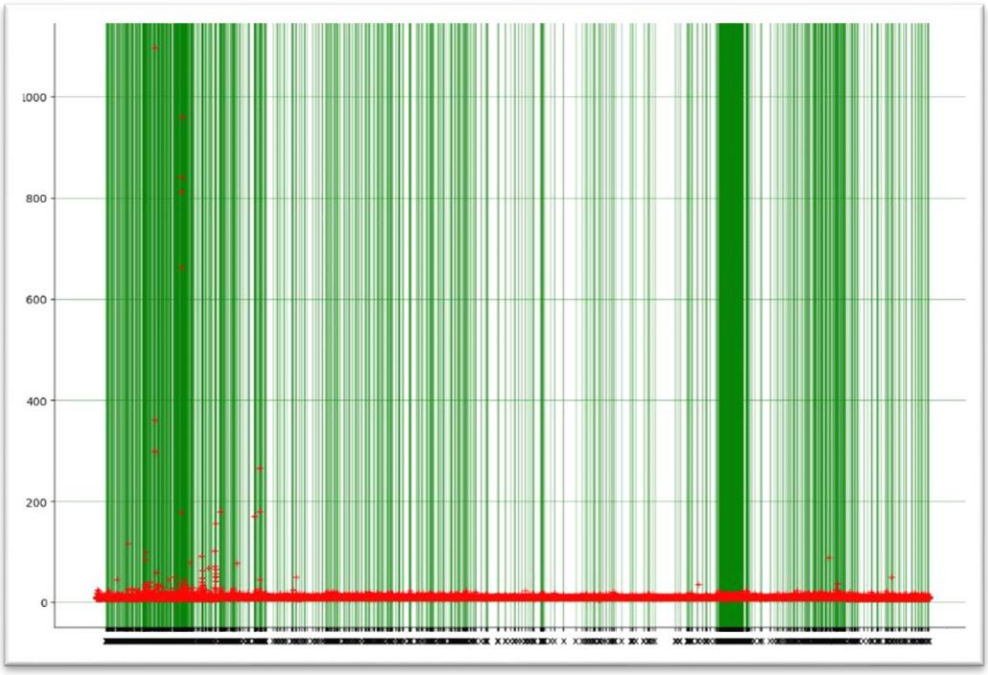
O código verifica se o valor de tempoDeResposta[i] está acima do limiteSuperior. Se isso ocorrer, significa que o valor é uma anomalia, e a variável numeroDeAnomalias é incrementada em 1. Além disso, o índice i é adicionado à lista instantes para registrar o instante em que a anomalia foi detectada.

- Exibição dos resultados:

Ao final do loop, o código imprime na tela o número total de anomalias detectadas e a porcentagem de anomalias em relação ao total de valores da matriz tempoDeResposta. Também são exibidos os índices dos instantes em que as anomalias foram detectadas.

O algoritmo retornou um total de 1321 anomalias detectadas (7.34% dos valores).

Gráfico 9: Plotagem de um gráfico com marcadores ("r+")



Fonte: (Bruscagini, 2023)

A imagem representa o gráfico de linha do conjunto de dados tempoDeResposta e adiciona marcadores de cruzes vermelhas ('r+') em posições específicas identificadas pelos índices contidos na lista instantes. O eixo x do gráfico será rotulado com a string 'x' em cada ponto onde os marcadores são exibidos, permitindo que esses pontos sejam facilmente identificados no gráfico. A grade verde é adicionada ao gráfico para facilitar a leitura e a análise dos dados.

3.6 Parte 6: Simplificando a detecção de anomalias

Embora a abordagem anterior possua uma sólida fundamentação teórica, sua aplicação prática revelou desafios significativos, não apenas em termos de eficácia, mas também em relação ao tempo dedicado à execução. A descoberta de um grande número de pontos rotulados como anomalias tornou-se inviável, dada a dificuldade que um administrador teria em investigar todos esses casos.

3.6.1 Avaliando uma abordagem mais simples

Diante das limitações identificadas na abordagem anterior, exploraremos uma alternativa mais simples que ainda possa ser eficaz na detecção de anomalias. Em

vez de utilizar análises complexas, consideraremos a média como um critério para avaliar quais pontos são considerados anomalias.

3.6.2 Utilizando a média como referência

Propomos uma abordagem direta: a partir da medida 2000, qualquer valor que exceda a média será identificado como uma anomalia. Esse método simplificado é menos oneroso computacionalmente e pode fornecer resultados mais acessíveis e de fácil interpretação.

3.6.3 Estabelecendo um ponto de referência adequado

Para assegurar que a média seja calculada com base em dados representativos, consideraremos o uso de valores coletados antes do ponto de referência (por exemplo, o primeiro cálculo da média). Isso nos permitirá contar com dados suficientes para o cálculo da média e, assim, estabelecer uma linha de corte adequada para a identificação de anomalias.

3.6.4 Contagem de valores anômalos

Uma vez que tenhamos definido a média como referência e estabelecido o ponto de início para o cálculo, contaremos quantos valores se enquadram na categoria de anomalias, ou seja, aqueles que excedem a média. Essa abordagem simplificada tem o potencial de fornecer uma maneira mais gerenciável de identificar e priorizar anomalias.

Ao refletirmos sobre a abordagem anterior, reconhecemos a necessidade de uma simplificação que equilibre eficácia e viabilidade. O uso direto da média como critério para a detecção de anomalias apresenta-se como uma alternativa mais acessível. Ao ajustar nosso foco para identificar valores que excedem a média, podemos obter resultados mais práticos e utilizáveis, proporcionando uma maneira eficiente de detectar comportamentos anômalos no servidor web.

Figura 16: Detecção de anomalias baseada na comparação entre cada valor atual da série temporal e a média das amostras

```
numeroDeAnomalias = 0
for i in range(199,len(tempoDeResposta)):
  if tempoDeResposta[i] > medias[i]:
    numeroDeAnomalias+=1
print(numeroDeAnomalias, "anomalias detectadas (%.2f%% dos valores)" %(100*numeroDeAnomalias/len(tempoDeResposta)))
```

Fonte: (Bruscagini, 2023)

A detecção de anomalias é feita comparando cada valor na posição i da matriz com a média das amostras anteriores, em um intervalo que varia de 199 até o final da matriz. A seguir uma explicação mais detalhada:

numeroDeAnomalias = 0: Inicializa a variável numeroDeAnomalias como 0. Essa variável será usada para contar o número de anomalias encontradas nos dados.

3.6.5 Loop para detecção de anomalias:

O código entra em um loop for i in range(199, len(tempoDeResposta)), que itera da posição 199 até o final da matriz tempoDeResposta. A posição 199 é usada como início para garantir que haja dados suficientes para calcular a média das amostras anteriores.

3.6.6 Comparação com a média:

Para cada iteração do loop, o código compara o valor em tempoDeResposta[i] com a média das amostras anteriores, medias[i]. A média é calculada utilizando as amostras de tempoDeResposta desde o início até a posição i, conforme implementado em algum trecho anterior do código que não foi fornecido.

3.6.7 Detecção de anomalias:

Se o valor em tempoDeResposta[i] for maior do que a média medias[i], isso indica que o valor é uma anomalia, e a variável numeroDeAnomalias é incrementada em 1.

3.6.8 Exibição dos resultados:

Ao final do loop, o código imprime na tela o número total de anomalias detectadas e a porcentagem de anomalias em relação ao total de valores da matriz tempoDeResposta.

O algoritmo calculou um total de 2640 anomalias detectadas (14.67% dos valores).

3.7 Parte 7: Refinando o critério de detecção de anomalias

A constatação de que considerar exatamente acima da média resultou em um excesso de anomalias nos leva a buscar um ajuste mais preciso. É essencial encontrarmos um equilíbrio. Nesse sentido, iremos refinar nosso critério de detecção de anomalias.

3.7.1 Ajuste necessário

A identificação estrita de valores logo acima da média não se mostrou eficaz, tendo em vista a maior quantidade de anomalias encontradas em comparação ao caso do boxplot. Reconhecemos, portanto, a necessidade de realizar um ajuste para alcançar resultados mais relevantes e gerenciáveis.

3.7.2 Estabelecendo um novo limiar

Propomos um novo critério para a detecção de anomalias. Dessa vez, consideraremos como anomalias os valores que forem pelo menos 10 vezes maiores que a média. Ao adotar esse limiar, buscamos obter um número significativamente reduzido de anomalias identificadas, o que nos permitirá focar nas mais relevantes.

3.7.3 Visualização gráfica aprimorada

Com o novo critério de detecção em vigor, procederemos à plotagem do gráfico original dos tempos de resposta. Destacaremos de forma explícita os pontos considerados como anomalias com base no limiar estabelecido. Essa visualização gráfica aprimorada proporcionará uma compreensão clara e imediata das áreas em que ocorrem as anomalias.

O processo de refinamento do critério de detecção de anomalias nos orientou a adotar uma abordagem mais precisa. Ao considerar valores substancialmente superior à média, visamos filtrar e priorizar as anomalias de maior relevância. Através da plotagem do gráfico original, destacando esses pontos, conseguiremos visualizar de maneira eficaz as regiões de maior interesse, permitindo uma análise mais concentrada e embasada.

Figura 17: Método de detecção de anomalias com comparação de cada valor i e limite superior definido.

```
numeroDeAnomalias = 0
instantes = []

for i in range(199,len(tempoDeResposta)):
  if tempoDeResposta[i] > 10*medias[i]:
    numeroDeAnomalias+=1
    instantes.append(i)
print(numeroDeAnomalias, "anomalias detectadas (%.2f%% dos valores)" %(100*numeroDeAnomalias/len(tempoDeResposta)))
print("Instantes com anomalias: ", instantes)
```

Fonte: (Bruscagini, 2023)

Este código é uma variação do método anterior de detecção de anomalias, mas com uma abordagem mais rigorosa. Ele busca identificar anomalias nos dados contidos na matriz tempoDeResposta ao comparar cada valor na posição i com um limite superior definido como 10 vezes a média das amostras anteriores, em um intervalo que varia de 199 até o final da matriz. A seguir maiores detalhes são apresentados:

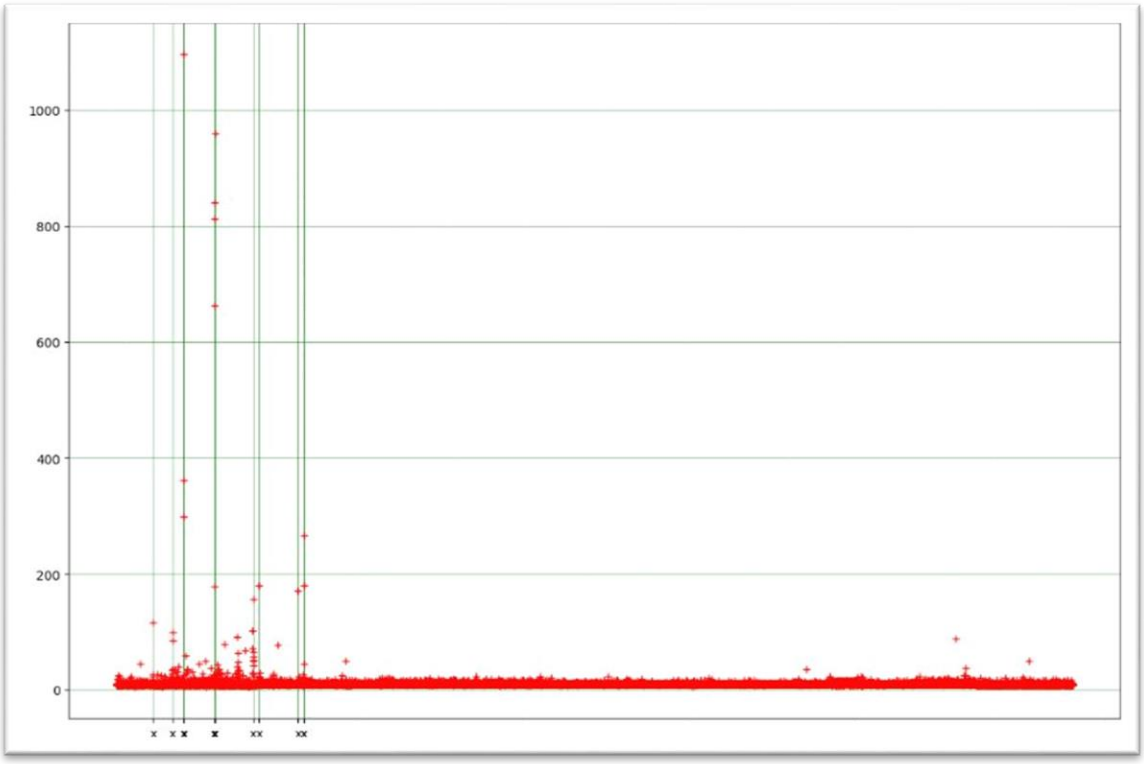
- numeroDeAnomalias = 0: Inicializa a variável numeroDeAnomalias como 0. Essa variável será usada para contar o número de anomalias encontradas nos dados.
  instantes = []: Inicializa a lista instantes, que será usada para armazenar os índices dos instantes em que as anomalias foram detectadas.
- Loop para detecção de anomalias:
  O código entra em um loop for i in range(199, len(tempoDeResposta)), que itera da posição 199 até o final da matriz tempoDeResposta. A posição 199 é usada como início para garantir que haja dados suficientes para calcular a média das amostras anteriores.
- Comparação com o limite superior:
  Para cada iteração do loop, o código compara o valor em tempoDeResposta[i] com o limite superior, definido como 10*medias[i], onde medias[i] é a média das amostras anteriores, calculada em algum trecho anterior do código que não foi fornecido.
- Detecção de anomalias:
  Se o valor em tempoDeResposta[i] for maior do que o limite superior (10*medias[i]), isso indica que o valor é uma anomalia. A variável

numeroDeAnomalias é incrementada em 1 e o índice i é adicionado à lista instantes para registrar o instante em que a anomalia foi detectada.

- Exibição dos resultados:
  Ao final do loop, o código imprime na tela o número total de anomalias detectadas e a porcentagem de anomalias em relação ao total de valores da matriz tempoDeResposta. Além disso, são exibidos os índices dos instantes em que as anomalias foram detectadas.

O algoritmo retornou um total de 15 anomalias detectadas (0.08% dos valores) e também os instantes das ocorrências: [686, 1050, 1262, 1263, 1264, 1830, 1831, 1832, 1846, 1847, 2564, 2677, 3409, 3524, 3526]

Gráfico 10: Representação do conjunto de dados “tempo de resposta” com marcadores



Fonte: (Bruscagini, 2023)

Este gráfico de linha representa o conjunto de dados tempoDeResposta e adiciona marcadores de cruzes vermelhas ('r+') em posições específicas identificadas pelos índices contidos na lista instantes. O eixo x do gráfico será rotulado com a string 'x' em cada ponto onde os marcadores são exibidos, permitindo que esses pontos sejam facilmente identificados no gráfico. A grade verde é adicionada ao gráfico para

facilitar a leitura e a análise dos dados. Essa abordagem é útil para destacar e visualizar pontos de interesse ou anomalias nos dados, facilitando a interpretação dos resultados e a identificação de comportamentos atípicos.

3.8 Parte 8: Explorando a análise de mudanças graduais

Uma alternativa adicional reside na análise das variações entre valores sucessivos, considerando a mudança em relação ao valor imediatamente anterior. Contudo, é importante adotar uma abordagem que seja sensível a mudanças significativas, mas sem exagerar nos critérios.

3.8.1 Avaliando mudanças graduais

A análise das variações entre valores adjacentes emerge como uma alternativa válida. Buscamos identificar aumentos ou diminuições suaves, alinhadas com a expectativa de que mudanças abruptas não devem ser uma ocorrência comum no sistema.

3.8.2 Evitando mudanças extremas

É essencial encontrar um ponto de equilíbrio ao definir a mudança significativa. Reconhecemos que aumentos de 10 vezes podem ser excessivos. Nossa meta é evitar valores muito altos, enquanto ainda capturamos mudanças que se desviam do padrão.

3.8.3 Um limiar sustentável

Ao estabelecer um limiar para mudanças significativas, consideraremos uma abordagem mais realista e sustentável. Nosso objetivo é detectar mudanças que representem uma diferença notável, mas que também sejam compatíveis com o comportamento esperado do sistema.

Ao explorar a análise de mudanças entre valores adjacentes, estamos adotando uma estratégia que busca um meio-termo entre a detecção sensível de variações e a moderação nos critérios. Ao evitar mudanças abruptas e excessivas, nossa abordagem visa identificar mudanças graduais que podem indicar possíveis anomalias no comportamento do servidor web.

Figura 18: Método de detecção de anomalias com comparação de cada valor i imediatamente anterior

```
numeroDeAnomalias = 0
instantes = []

for i in range(199,len(tempoDeResposta)):
  # Em alguns casos, pode ser interessante verificar o valor absoluto da diferença
  if tempoDeResposta[i] > 10*tempoDeResposta[i-1]:
    numeroDeAnomalias+=1
    instantes.append(i)
print(numeroDeAnomalias, "anomalias detectadas (%.2f%% dos valores)" %(100*numeroDeAnomalias/len(tempoDeResposta)))
print("Instantes com anomalias: ", instantes)
```

Fonte: (Bruscagini, 2023)

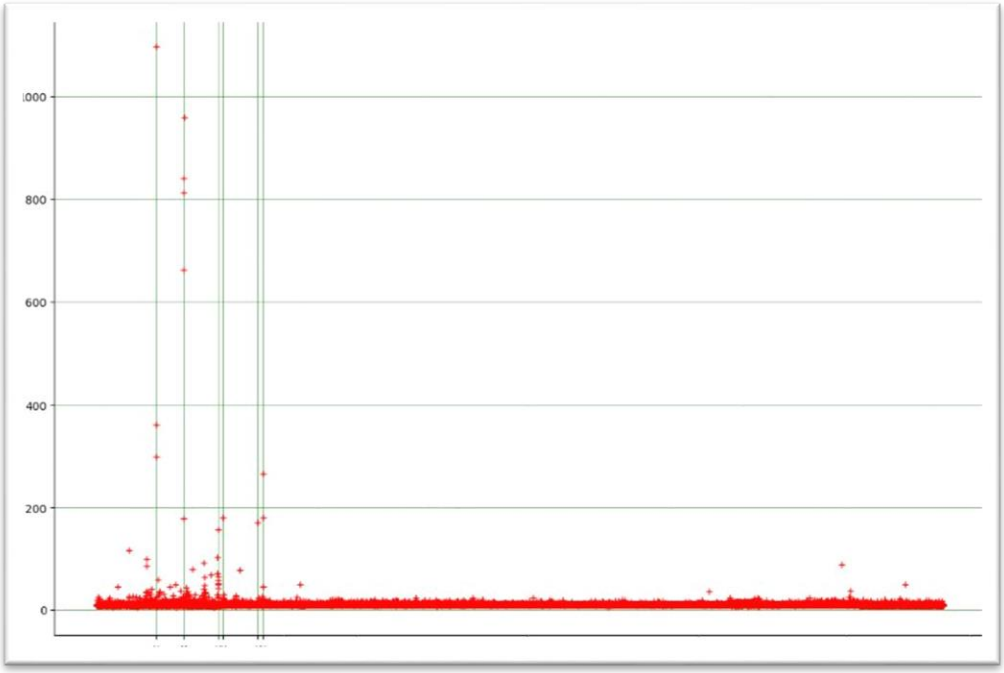
Este código é uma variação do método anterior de detecção de anomalias, mas com uma abordagem diferente. Neste caso, a detecção de anomalias é realizada comparando cada valor na posição i com o valor imediatamente anterior (tempoDeResposta[i-1]), em vez de usar a média das amostras anteriores. A seguir maiores detalhes:

- numeroDeAnomalias = 0: Inicializa a variável numeroDeAnomalias como 0. Essa variável será usada para contar o número de anomalias encontradas nos dados.
- instantes = []: Inicializa a lista instantes, que será usada para armazenar os índices dos instantes em que as anomalias foram detectadas.
- Loop para detecção de anomalias:
  O código entra em um loop for i in range(199, len(tempoDeResposta)), que itera da posição 199 até o final da matriz tempoDeResposta. A posição 199 é usada como início para garantir que haja dados suficientes para comparar o valor atual com o valor anterior.
- Comparação com o valor anterior:
  Para cada iteração do loop, o código compara o valor em tempoDeResposta[i] com o valor imediatamente anterior tempoDeResposta[i-1].
- Detecção de anomalias:
  Se o valor em tempoDeResposta[i] for maior do que 10 vezes o valor anterior (tempoDeResposta[i-1]), isso indica que o valor é uma anomalia. A variável numeroDeAnomalias é incrementada em 1 e o índice i é adicionado à lista instantes para registrar o instante em que a anomalia foi detectada.
- Exibição dos resultados:
  Ao final do loop, o código imprime na tela o número total de anomalias detectadas e a porcentagem de anomalias em relação ao total de valores da matriz tempoDeResposta. Além disso, são exibidos os índices dos instantes em que as anomalias foram detectadas.

Essa abordagem de detecção de anomalias compara o valor atual com o valor imediatamente anterior, permitindo identificar alterações bruscas nos dados que possam indicar comportamentos anômalos. O uso do valor absoluto da diferença (abs(tempoDeResposta[i] - tempoDeResposta[i-1])) pode ser interessante para capturar tanto anomalias positivas quanto negativas, caso sejam relevantes. A escolha do fator de multiplicação (10 neste caso) pode ser ajustada de acordo com o contexto e a sensibilidade desejada para a detecção de anomalias.

Aqui o algoritmo retornou 7 anomalias detectadas (0.04% dos valores) e os respectivos instantes das ocorrências: [1262, 1830, 1846, 2564, 2677, 3409, 3524].

Gráfico 11: Representação do conjunto de dados “tempo de resposta” para o novo método.



Fonte: (Bruscagini, 2023)

O gráfico de linha representa o conjunto de dados tempoDeResposta e adiciona marcadores de cruzes vermelhas ('r+') em posições específicas identificadas pelos índices contidos na lista instantes. O eixo x do gráfico será rotulado com a string 'x' em cada ponto onde os marcadores são exibidos, permitindo que esses pontos sejam facilmente identificados no gráfico. A grade verde é adicionada ao gráfico para facilitar a leitura e a análise dos dados. Essa abordagem é útil para destacar e visualizar pontos de interesse ou anomalias nos dados, facilitando a interpretação dos resultados e a identificação de comportamentos atípicos.

4 ANÁLISE DOS RESULTADOS

Este trabalho visa fornecer uma análise detalhada dos resultados obtidos por meio da aplicação de diversos métodos de detecção de anomalias em um conjunto de dados de tempo de resposta. Cada método explorado oferece uma abordagem única para identificar comportamentos atípicos nos dados, contribuindo para uma compreensão mais profunda das variações e padrões presentes.

4.1 Método das Médias Móveis Cumulativas: Explorando Tendências Temporais

O "Método das Médias Móveis Cumulativas" foi aplicado com sucesso neste estudo. Essa abordagem envolve o cálculo das médias móveis ao longo de intervalos crescentes de dados. O objetivo é suavizar as flutuações e destacar tendências subjacentes. O gráfico resultante exibe uma curva que enfatiza a evolução ao longo do tempo, permitindo a identificação de padrões de longo prazo e facilitando a interpretação dos dados. Esse método é particularmente útil para analisar o comportamento geral dos dados, revelando tendências que podem não ser evidentes em uma visualização simples.

4.2 Método dos Múltiplos Gráficos de Médias Móveis: Uma Abordagem Multiescalar

A abordagem dos "Múltiplos Gráficos de Médias Móveis" foi utilizada para explorar as variações dos dados em diferentes intervalos de amostragem. Ao gerar vários gráficos, cada um com um intervalo diferente, pudemos analisar as tendências em várias escalas temporais. Essa técnica é valiosa para identificar padrões que podem ser obscurecidos quando observados em uma única escala. A análise multiescalar oferece uma perspectiva mais completa da dinâmica dos dados e ajuda a revelar insights ocultos.

4.3 Explorando a Distribuição por Meio de Boxplots: Análise Visual de Discrepâncias

A análise de boxplots é uma ferramenta poderosa para compreender a distribuição dos dados. O "Boxplot Geral" proporciona uma visão abrangente da dispersão dos dados em relação aos quartis e possíveis outliers. Isso auxilia na

identificação de valores extremos que podem indicar situações anômalas ou comportamentos incomuns. Já o "Boxplot Intervalado" nos permite focar em uma seção específica dos dados, fornecendo uma análise mais detalhada da distribuição nesse intervalo. Ambos os métodos oferecem insights sobre a variabilidade dos dados e auxiliam na identificação de padrões.

4.4 Métodos de Detecção Baseados em Limites e Comparações: Identificação de Anomalias Pontuais

Diversos métodos de detecção de anomalias baseados em limites e comparações foram explorados. O "Método para Detectar Anomalias Utilizando o Conceito de Intervalo Interquartil (IQR)" é uma abordagem estatística que se baseia na diferença entre o terceiro e o primeiro quartil. Ele identifica valores que se afastam significativamente da distribuição central dos dados. Por outro lado, o "Método de Detecção de Anomalias Comparando com o Valor Imediatamente Anterior" procura alterações bruscas entre valores consecutivos. Ambos os métodos oferecem uma abordagem pontual para identificar comportamentos atípicos e são particularmente úteis para identificar anomalias que se destacam em relação ao contexto imediato.

4.5 Comparação de Métodos de Detecção de Anomalias

Este relatório apresenta uma análise abrangente da eficácia de diferentes métodos de detecção de anomalias aplicados a um conjunto de dados de tempo de resposta. Os métodos foram avaliados com base em sua capacidade de identificar anomalias de maneira precisa e eficiente, contribuindo para insights significativos na análise de dados.

Boxplots (Geral e Intervalado): A análise por meio de boxplots se destaca como o método mais eficaz para detectar anomalias, oferecendo uma visão visual clara das distribuições estatísticas dos dados. O boxplot geral revela a amplitude e quartis da distribuição, enquanto o boxplot intervalado permite focar em segmentos específicos dos dados. Essa abordagem proporciona insights detalhados sobre a variabilidade e a presença de outliers, sendo altamente eficaz na detecção de anomalias.

Método das Médias Móveis Cumulativas: Esse método oferece uma compreensão profunda das tendências temporais dos dados, destacando padrões de longo prazo. A visualização das médias móveis suaviza as flutuações e ressalta a evolução ao longo do tempo. Embora não seja tão sensível para a detecção de anomalias pontuais, é altamente eficaz na identificação de padrões de comportamento.

Múltiplos Gráficos de Médias Móveis: Ao explorar os dados em diferentes escalas temporais, essa abordagem é valiosa para identificar padrões em várias resoluções. No entanto, sua eficácia pode variar dependendo do intervalo escolhido. Em algumas situações, essa abordagem pode capturar tendências de maneira mais precisa do que outros métodos, mas pode ser menos sensível para detectar anomalias sutis.

Método de Detecção Baseado em Limites e Comparações: Embora os métodos de detecção baseados em limites, como o conceito de intervalo interquartil (IQR), e aqueles que comparam com valores imediatamente anteriores possam identificar anomalias, eles tendem a ser menos sensíveis a variações sutis e podem não ser ideais para detectar comportamentos atípicos menos óbvios.

5 CONCLUSÃO

A segurança cibernética emerge como uma preocupação incontornável nesse ambiente digital altamente interconectado. A crescente quantidade de dados gerados, armazenados e transmitidos exige uma abordagem sólida para proteger essas informações contra uma miríade de ameaças cibernéticas, que se tornaram cada vez mais sofisticadas e perniciosas. As técnicas de análise de dados desempenham um papel vital nesse empreendimento, permitindo que as organizações identifiquem padrões suspeitos, comportamentos anômalos e indicadores de ataques potenciais.

A análise de dados se manifesta como uma ferramenta proativa para a detecção precoce de intrusões, identificação de vulnerabilidades e monitoramento de atividades suspeitas. Através de abordagens estatísticas e algoritmos avançados, é possível não apenas reagir a incidentes de segurança, mas também prevenir e mitigar possíveis ameaças. Ao empregar técnicas de análise, como análise de séries temporais, detecção de anomalias e análise comportamental, as organizações podem identificar atividades não usuais que podem indicar uma violação de segurança em estágios iniciais.

A abordagem integrada de métodos analíticos discutida neste estudo ressalta sua relevância crítica para a detecção de anomalias e sua eficácia na identificação de eventos anômalos, fornecendo insights valiosos para as equipes responsáveis.

A detecção de anomalias é uma tarefa de suma importância em uma variedade de contextos, desde monitoramento de sistemas críticos até análise de dados em campos diversos, como medicina, finanças e engenharia. A capacidade de identificar e isolar essas anomalias é fundamental para garantir a integridade, a segurança e o desempenho de processos e sistemas.

Nesse sentido, a escolha cuidadosa e adaptativa dos métodos de análise se torna ainda mais crucial. A seleção do método deve ser orientada pelo contexto específico e pelos objetivos da análise, pois isso influenciará diretamente na sensibilidade na detecção de anomalias. Equilibrar essa sensibilidade com a minimização de falsos positivos é um desafio contínuo, mas garantir resultados confiáveis e acionáveis é a recompensa.

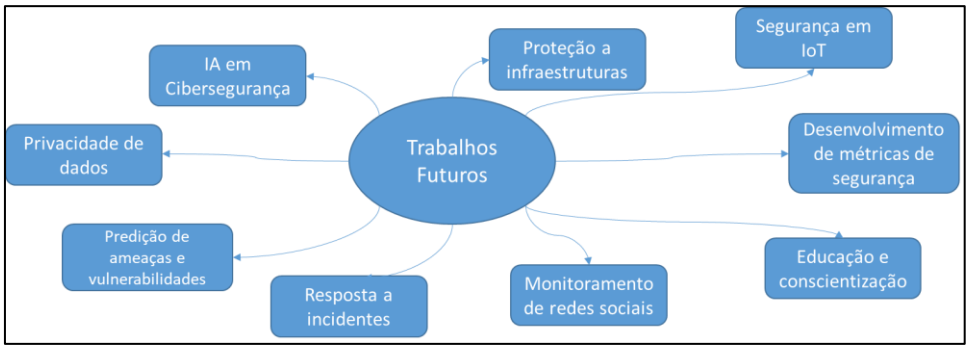
A análise comparativa dos métodos ressalta que, ao escolher a abordagem certa, as equipes responsáveis podem efetivamente destacar distribuições, tendências temporais e outliers que, de outra forma, poderiam passar despercebidos. Boxplots, com sua capacidade de destacar distribuições e outliers, são altamente recomendados para obter uma visão geral. Quando a ênfase recai sobre a identificação de tendências temporais, o método das médias móveis cumulativas se destaca. Por outro lado, para explorar resoluções temporais variadas, os múltiplos gráficos de médias móveis se mostram uma opção valiosa.

Para análises futuras, a integração de múltiplos métodos continua a ser uma estratégia promissora para aumentar a sensibilidade na detecção de anomalias, proporcionando à equipe uma vantagem significativa na identificação precoce de eventos adversos. Além disso, a exploração de métodos avançados, especialmente aqueles baseados em aprendizado de máquina, pode ser benéfica ao lidar com cenários complexos e padrões não lineares. Em resumo, a escolha do método certo e a sua adaptação ao contexto e aos objetivos específicos da análise são essenciais para garantir que as equipes responsáveis estejam equipadas com as ferramentas mais eficazes na detecção de anomalias, proporcionando um ambiente mais seguro, confiável e eficiente em suas operações.

6 TRABALHOS FUTUROS

À medida que a sociedade continua a avançar em direção a uma era cada vez mais digital e interconectada, a análise de dados e a segurança cibernética emergem como campos cruciais para garantir a integridade, confidencialidade e disponibilidade das informações.

Figura 19 - Representação dos tópicos de trabalhos futuros.



Fonte: (Bruscagini, 2023)

No entanto, à medida que enfrentamos desafios crescentes e complexos no mundo digital, novas oportunidades de pesquisa e inovação surgem para abordar as lacunas existentes e enfrentar ameaças em constante evolução. Abaixo estão algumas áreas potenciais de interesse para futuros trabalhos, considerando os temas tratados nesta conversa.

6.1 Integração de Inteligência Artificial na Cibersegurança

A aplicação de técnicas de inteligência artificial (IA) na segurança cibernética é um campo em crescimento. Explorar como algoritmos de IA podem detectar e responder automaticamente a ameaças em tempo real, aprender com padrões de comportamento e antecipar ataques desconhecidos oferece um vasto terreno para pesquisa. Além disso, a combinação de IA e análise de dados pode resultar em sistemas de defesa cibernética mais eficazes e adaptativos.

6.2 Privacidade de Dados e Ética em Análise de Dados

À medida que coletamos e analisamos dados para fins de segurança, é vital considerar questões de privacidade e ética. Explorar como proteger as informações pessoais dos usuários enquanto se obtêm insights úteis para a segurança é um

desafio fundamental. Pesquisas podem se concentrar em métodos para garantir a anonimização dos dados, desenvolver políticas de privacidade robustas e abordar preocupações éticas relacionadas à coleta e uso de dados.

6.3 Análise de Dados em Tempo Real

Com a proliferação de dispositivos conectados e o aumento da velocidade das comunicações, a análise de dados em tempo real torna-se crucial para identificar ameaças em estágios iniciais. Pesquisas futuras podem se concentrar em desenvolver algoritmos de análise de dados que possam processar informações em tempo real, detectando anomalias e atividades suspeitas com eficácia e rapidez.

6.4 Predição de Ameaças e Vulnerabilidades

A previsão de ameaças cibernéticas e vulnerabilidades é uma área em crescimento. Utilizando técnicas de análise de dados e aprendizado de máquina, os pesquisadores podem explorar como antecipar possíveis ataques e identificar vulnerabilidades antes que sejam exploradas. Isso permitiria às organizações tomar medidas proativas para mitigar riscos e fortalecer suas defesas.

6.5 Segurança em Sistemas de Internet das Coisas

Com a crescente adoção de dispositivos de Internet das Coisas (IoT), surgem desafios únicos em relação à segurança. Futuras pesquisas podem explorar como a análise de dados pode ser aplicada para proteger redes e dispositivos IoT contra ataques, garantindo a integridade das comunicações e a privacidade dos usuários.

6.6 Desenvolvimento de Métricas de Avaliação de Segurança

A avaliação da eficácia das estratégias de segurança é fundamental. Futuros trabalhos podem se concentrar no desenvolvimento de métricas de avaliação de segurança que permitam às organizações medir a eficácia de suas defesas cibernéticas. Isso envolveria a análise de dados para identificar padrões de ataque e determinar a eficácia das medidas de proteção.

6.7 Educação e Conscientização em Segurança Cibernética

A conscientização e a educação dos usuários desempenham um papel crucial na segurança cibernética. Pesquisas futuras podem se concentrar em desenvolver

métodos inovadores de educação e treinamento que ajudem os usuários a reconhecer ameaças e adotar práticas seguras online.

6.8 Monitoramento de Redes Sociais para Detecção de Ameaças

As redes sociais podem ser utilizadas para disseminar ameaças e ataques cibernéticos. Explorar como a análise de dados pode ser aplicada para monitorar as redes sociais em busca de atividades maliciosas e identificar ameaças emergentes é uma área promissora.

6.9 Resposta a Incidentes com Base em Dados

A análise de dados pode desempenhar um papel fundamental na resposta a incidentes cibernéticos. Pesquisas futuras podem investigar como a análise de dados pode agilizar a identificação da origem e do impacto de um incidente, facilitando a recuperação e minimizando danos.

6.10 Proteção de Infraestruturas Críticas

As infraestruturas críticas, como redes elétricas e sistemas de transporte, são alvos potenciais de ataques cibernéticos. Pesquisas futuras podem explorar como a análise de dados pode ser aplicada para monitorar e proteger essas infraestruturas, garantindo sua operação segura e confiável.

Entretanto, as áreas de análise de dados e segurança cibernética estão intrinsecamente interligadas, oferecendo um vasto leque de oportunidades para pesquisa e inovação. Nesse sentido, a busca por soluções inovadoras e sustentáveis em várias áreas do conhecimento está impulsionando o avanço da pesquisa interdisciplinar.

Três áreas de pesquisa específicas estão emergindo como pilares de desenvolvimento: bioinformática, geração de energia nuclear e inteligência artificial. Cada uma dessas áreas oferece desafios e oportunidades únicas, contribuindo para a melhoria da saúde humana, a evolução da indústria energética e a capacidade de sistemas de aprendizado profundo.

- Bioinformática: Explorando o Potencial da Biologia Computacional

A bioinformática está revolucionando a maneira como compreendemos e exploramos o mundo biológico. Ao combinar biologia e computação, essa área de pesquisa permite análises detalhadas de sequências genéticas, modelagem de proteínas e simulação de processos biológicos complexos. Através da bioinformática, os pesquisadores estão identificando novos alvos para medicamentos, melhorando a previsão de doenças e avançando na medicina personalizada. O estudo de vias metabólicas, análise de dados genômicos e a aplicação de algoritmos de aprendizado de máquina estão transformando nossa compreensão da vida e acelerando a descoberta de soluções para doenças e desafios globais.

- Geração de Energia Nuclear: Moldando o Futuro da Energia Limpa

A geração de energia nuclear está evoluindo para enfrentar as demandas crescentes por fontes de energia limpa e confiável. Reatores avançados e tecnologias de geração IV estão sendo desenvolvidos para melhorar a segurança e eficiência dos sistemas nucleares. Além disso, a pesquisa em reatores de pequeno porte e modulares visa oferecer soluções flexíveis e descentralizadas. A fusão nuclear, ainda em estágios experimentais, promete uma fonte quase inesgotável de energia. Ao mesmo tempo, abordagens de segurança avançadas e a integração com fontes renováveis estão definindo o futuro da energia nuclear. Essa pesquisa é essencial para garantir um suprimento energético sustentável enquanto reduzimos as emissões de carbono.

- Inteligência Artificial/Aprendizado de Máquina

A inteligência artificial está revolucionando diversas indústrias, impulsionada pelo poder do aprendizado de máquina e do processamento de dados em larga escala. A pesquisa em IA está focada em melhorar a capacidade de sistemas de aprendizado profundo entenderem e responderem a dados complexos. Isso tem aplicações em campos como diagnóstico médico, previsão climática, processamento de linguagem natural e tomada de decisões autônomas. A interseção da IA com a bioinformática e a geração de energia nuclear oferece oportunidades para otimizar processos, acelerar descobertas científicas e melhorar a eficiência operacional.

À medida que a pesquisa nessas áreas avança, a interdisciplinaridade se torna cada vez mais crucial. A colaboração entre cientistas, engenheiros, matemáticos, médicos e especialistas de várias áreas é fundamental para enfrentar os desafios

complexos que essas áreas de pesquisa apresentam. Através de uma abordagem integrada, podemos explorar novas fronteiras do conhecimento, aproveitar o potencial transformador da tecnologia e moldar um futuro mais promissor e sustentável.

REFERÊNCIAS

(s.d.).

ABNT NBR. (2013). *ABNT NBR ISO 27001: Sistemas de Gestão de Segurança da Informação*. Fonte: ABNT: https://www.normas.com.br/visualizar/abnt-nbr-nm/25074/nbriso-iec27001-seguranca-da-informacao-seguranca-cibernetica-e-protecao-a-privacidade-sistemas-de-gestao-da-seguranca-da-informacao-requisitos#:~:text=Este%20documento%20especifica%20os%20requisitos%

Accenture. (2021). *Cyber Threatscape Report 2021.*

Ahad, M. A. (2019). Machine learning and deep learning approaches for cybersecurity: A survey. *IEEE Access*.

ANSC. (2023). *Agência Nacional de Segurança Cibernética (ANSC)*. Fonte: https://www.ansc.gov.br/

Antonakakis, M. P.-N. (2010). From throw-away traffic to bots: Detecting the rise of DGA-based malware. . *Proceedings of the 21st USENIX conference on Security Symposium*.

Brasil. (2014). Fonte: LEI Nº 12.965, DE 23 DE ABRIL DE 2014.

Brasil. (2018). *LEI Nº 13.709, DE 14 DE AGOSTO DE 2018*.

Brockwell, P. J. (2016). Introduction to time series and forecasting. *Springer*.

BRONZATTI, A. M. (2021). Os atores não-estatais e estatais no mundo cibernético e sua influência no sistema internacional. .

BRUSCAGINI, L. e. (2021). Data Security, Privacy, and Regulatory Issues: A Conceptual Approach to Digital Transformation to Smart Cities. . *Brazilian Technology Symposium. Cham: Springer International Publishing*, pp. p. 256-263.

BRUSCAGINI, L. e. (2021). Mathematical modeling: a conceptual approach of linear algebra as a tool for technological applications. *Brazilian Technology Symposium. Springer International Publishing*, pp. p. 239-248.

CARDOZO, R. L. (2019). A inclusão na base da pirâmide: um estudo brasileiro de aspectos motivadores e inibidores da adoção da tecnologia mobile banking pelo público de baixa renda.

CASTELLS, M. (2020). *Fim de milênio-A Era da Informação-vol. 3.* Editora Paz e Terra.

Cert.br. (s.d.). *Incidentes Notificados ao CERT.br*. Fonte: https://stats.cert.br/incidentes/

Chandola, V. B. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*.

CROWDSTRIKE. (29 de junho de 2023). *CrowdStrike Global Threat Report 2023.* Fonte: https://www.crowdstrike.com/resources/infographics/global-threat-report-2023/

EC-Council. (2023). *Ethical Hacking Essentials Copyright*. Fonte: EC-Council: https://codered.eccouncil.org/course/ethical-hacking-essentials?logged=false

Eskin, E. A. (2002). A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data. . *Applications of data mining in computer security*, pp. 77-101.

EXAME. (29 de junho de 2023). *58% dos brasileiros sofreram crimes cibernéticos, aponta estudo da Norton. .* Fonte: EXAME: https://exame.com/tecnologia/58-dos-brasileiros-sofreram-crimes-ciberneticos-aponta-estudo-da-norton/

EXPERIAN, S. (2021). *Ano de 2021 bate recorde com mais de 4 milhões de tentativas de fraude*. Fonte: https://www.serasaexperian.com.br/sala-de-imprensa/analise-de-dados/ano-de-2021-bate-recorde-com-mais-de-4-milhoes-de-tentativas-de-fraude-revela-serasa-experian/

FERREIRA, A. R. (2017). ISO/IEC 27001:2013 information security management system implementation in a university: A case study. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*.

FTC. (2019). *Equifax Data Breach Settlement.* Fonte: https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement

Gao, J. D. (2017). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*.

GDPR. (2018). *General Data Protection Regulation.* Fonte: European Union. : https://gdpr.eu/

Han, B. K. (2020). Machine learning in support of internet of things: A survey. . *IEEE Internet of Things Journal*.

HIPAA. (2021). *Health Information Privacy.* . Fonte: U.S. Department of Health & Human Services.: https://www.hhs.gov/hipaa/index.html

HOSANG, A. (2011). Política Nacional de Segurança Cibernética: uma necessidade para o Brasil.

ISO/IEC. (2021). *ISO/IEC 27001 Information Security Management.* Fonte: International Organization for Standardization. : https://www.iso.org/isoiec-27001-information-security.html

JUSTIÇA, M. D. (29 de junho de 2023). *Dados Nacionais de Segurança Pública.* Fonte: https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/estatistica/dados-nacionais-1/dados-nacionais

LESSA, D. T. (2021). Cartografias dos movimentos sociais nos livros didáticos de geografia. .

Lippi, M. &. (2018). Context-based routing for advanced driver assistance systems. . *IEEE Transactions on Intelligent Transportation Systems*.

MEIRELLES, F. S. (2022). *Pesquisa do Uso da TI-Tecnologia de Informação nas Empresas.* FGV EAESP.

MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA. (2021). Fonte: DADOS E INFORMAÇÕES NACIONAIS DE SEGURANÇA PÚBLICA: https://www.gov.br/mj/pt-br/assuntos/sua-seguranca/seguranca-publica/estatistica/dados-nacionais-1/dados-nacionais

Mirkovic, J. &. (2017). A Taxonomy of DDoS Attack and DDoS Defense Mechanisms. . *ACM Computing Surveys (CSUR)*.

NAGARAJ, K., SHARVANI, G., & SRIDHAR, A. (2018). Emerging trend of big data analytics in bioinformatics: a literature review. *International Journal of Bioinformatics Research and Applications*, pp. n. 1-2, p. 144–205.

NIST. (2023). Fonte: https://csrc.nist.gov/publications/detail/sp/800-95/rev-1/final

OWASP. (2021). *Cheat Sheet. Authentication cheat sheet.*

OWASP. (s.d.). *OWASP Top Ten*. Acesso em 2023, disponível em https://owasp.org/www-project-top-ten/

Pancini, L. (2022). *58% dos brasileiros sofreram crimes cibernéticos, aponta estudo da Norton*. Fonte: Exame: https://exame.com/tecnologia/58-dos-brasileiros-sofreram-crimes-ciberneticos-aponta-estudo-da-norton/

PSI. (2023). *https://www.pcisecuritystandards.org/about_us/*.

Registry of Open Data on AWS. (2023). *A Realistic Cyber Defense Dataset (CSE-CIC-IDS2018)*. Fonte: AWS: https://registry.opendata.aws/cse-cic-ids2018/

RON, D. (2016). Briefing-big data and data analytics-the potential for innovation and growth. *EASO: European Asylum Support Office*.

SANS. (2010). *Intrusion Detection. Can you explain traffic analysis and anomaly detection. .*

Security Configuration Checklists Program for IT Products: Guidelines for Web Servers. Fonte: NIST: . (2021). Acesso em 2023, disponível em https://csrc.nist.gov/publications/detail/sp/800-123/rev-2/final

STEVENSON, D. (1997). Information and Communications Technology in UK Schools. An independente inquiry. . *The Independent ICT in Schools Commission. .*

STORTI, J. M. (2009). Enfrentado as novas ameaças: Estratégia e política internacional norte-americanas no pós-Guerra Fria.

VERIZON, B. (2008). *2008 Data Breach Investigations Report.*